The Dissertation Committee for Garrett Patrick Hagemann
certifies that this is the approved version of the following dissertation:

Essays in Nonlinear Pricing

Committee:

Eugenio Miravete, Supervisor

Robert Town

Daniel Ackerberg

Katja Seim

Essays in Nonlinear Pricing

by

Garrett Patrick Hagemann

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2018

To Mary: your unwavering support made this possible.

Acknowledgments

I would like to thank Eugenio Miravete, Robert Town, Daniel Ackerberg, and Katja Seim for their invaluable guidance and comments.

Essays in Nonlinear Pricing

Publication No. ____________________

Garrett Patrick Hagemann, Ph.D.
The University of Texas at Austin, 2018

Supervisor: Eugenio Miravete

This dissertation addresses several open issues in the economics surrounding the use of nonlinear pricing. The first chapter empirically examines the impact of the use of nonlinear pricing by wholesalers. The second chapter evaluates how firm profit depends on the number of prices offered in a nonlinear price schedule. Finally, the third chapter investigates the use of all-unit discounts as a price discrimination instrument.

The first chapter exploits a unique data set of price schedules to provide the first empirical estimate of the welfare impact of second degree price discrimination in a market with double marginalization. Theoretical predictions in such a context are ambiguous. Quantity discounts at the wholesale level reduce costs for larger retailers, increasing efficiency. However, quantity discounts raise input costs for smaller retailers, increasing prices consumers may pay. The combined welfare effects on consumers depends on how much of input cost discounts are passed through to consumers and the distribution of retailer size. I develop and estimate a model of the New York State retail liquor market where wholesalers offer a multi-part nonlinear tariff for each product. The structural model is then used to estimate

the welfare impact of restricting wholesale pricing to be linear. I find that banning quantity discounts reduces total welfare by approximately 14% on average. Consumer surplus and wholesaler profit decline by approximately 26% on average. Average retailer profit increases by a similar magnitude, though effects for a particular retailer are heterogeneous across retailer size.

The second chapter examines the shape of observed price schedules more directly. Sellers often offer price schedules with relatively few segments rather than completely nonlinear price schedules which offer a unique price for each unit sold. By not offering a completely nonlinear, sellers are foregoing some additional profit in favor of a simpler pricing strategy. I find that the scale of these foregone profits is relatively small and only loosely related to product characteristics. When considered in percentage terms, foregone profits are very similar across a large number of products. This suggests that simple pricing strategies obtain almost all the profits available and this is a common property of nonlinear pricing strategies.

The final chapter compares price discrimination through two different quantity discount mechanisms: all-unit discounts and incremental discounts. All-unit-discounts give consumers a lower marginal price on all units purchased once total purchase size crosses a threshold. Incremental discounts only provide discounts on units above the threshold. Relative to incremental discounts, all-unit-discounts imply higher marginal prices and bunching of purchase sizes in equilibrium. The equilibrium bunching may present a challenge for estimating the model empirically.

Table of Contents

List of Tables

List of Figures

Chapter 1

Upstream Quantity Discounts and Double Marginalization in the New York Liquor Market

1.1 Introduction

The welfare impacts of second degree price discrimination[1] in an oligopoly setting are not well understood. This is true in both final goods markets as well as in the case I study, intermediate input markets. For intermediate goods markets, the welfare conclusions depend on the responsiveness of the final demand to price, the nature of competition in the final goods segment, and on the degree of heterogeneity of firms using the intermediate good as an input.[2] Given that the theoretical predictions hinge on these market characteristics, the impact of allowing or not allowing second degree price discrimination must be assessed empirically. However, lack of sufficiently detailed data on nonlinear tariffs has precluded much empirical work on second degree price discrimination in intermediate goods markets. Thus, little is known about the implications of such nonlinear pricing strategies in these markets. This paper seeks to address this gap.

[1]Second degree price discrimination is defined as a mechanism whereby consumers self select among a menu of prices. This is distinct from first degree price discrimination, where the seller charges exactly the buyer's willingness to pay, and third degree price discrimination, where the seller uses an observable characteristic of the buyer to charge different prices to different groups. See (Tirole 1988) §3 for more discussion of various forms of price discrimination.

[2]Appendix A provides a simple numerical example exploring these issues further.

Specifically, I evaluate the welfare consequences of wholesale quantity discounts in an intermediate goods market. Quantity discounts, where higher purchase volumes result in a lower per unit price, are a common form of second degree price discrimination. The setting I examine is the wholesale distribution of liquor in New York State. The state liquor regulator allows wholesalers to set nonlinear price schedules in quantities. Retailers who purchase a single case of liquor will pay more per bottle than retailers who purchase more cases of the same liquor from the same wholesaler. Retailers are heterogeneous because of income or other demographic attributes of their customers. Downstream heterogeneity on willingness to pay allows wholesalers to increase profit by offering these menus of tariffs.

As mentioned, one of the difficulties in examining the impact of upstream nonlinear pricing is the lack of nonlinear tariff information. I overcome that challenge by collecting a data set containing detailed pricing data from New York State. Liquor wholesalers in New York State operate under a "post and hold" system where they are mandated by law to report prices for the following month at the beginning of each month for every product they offer. I collect these reported price schedules from every month for 2009 through 2013. More than 2,000 price schedules are reported monthly. This data set is unique in that it details price schedules for many different products over a long time horizon. I match these price schedules to downstream retail purchases from the Nielsen Homescan Panel. The Nielsen data provides individual level purchase decisions and demographic characteristics. The retail purchase information covers the same time span as the price schedule data.

In this paper, I develop a structural model of wholesaler second degree price discrimination and exploit the rich data available on price schedules to estimate the supply side parameters including wholesaler marginal cost and the distribution of retailer heterogeneity.

The model consists of three components. First, wholesalers that set a quantity discount schedule for each product. Second, retailers that take the wholesale price schedules as given and set prices faced by consumers. Finally, consumers that observe retail prices and make purchase decisions.

Rather than offering a completely nonlinear schedule, where each case is offered at a unique price, wholesalers offer multi-part tariffs. These multi-part tariffs are a series of price and quantity bands where price per unit is piece-wise constant within that quantity band. This feature is common to all the price schedules described in my data. These multi-part tariffs are common across industries; examples can be readily found in industrial supplies,[3] office supplies,[4] and the financial sector.[5]

An innovation of this paper is to consider these multi-part tariffs as an equilibrium object and incorporate the optimal design of the tariffs into the approach used to recover the parameters of the structural model. The number of price segments, the prices themselves, and the quantity cutoffs all vary across products. Likewise, there is variation within a product across time. Further, this variation does not appear to be strongly related to product characteristics such as spirit type, import status, or bottle size.

I implement an approach that exploits the detailed nature of the price schedule data and recovers the supply side parameters from the shape of the price schedules. My approach relies on moment inequalities implied by the profit maximization choices of the wholesaler. I recover the parameters of the structural model by comparing profits under the observed

[3]www.fastenersuperstore.com/ offers multi-carton discounts for a wide array of fasteners and screws.
[4]www.uline.com offers per unit discounts on many common office, janitorial, and safety supplies.
[5](Copeland and Garratt 2015) detail a nonlinear price schedule for settling financial accounts.

price schedule with perturbations of that price schedule. I evaluate the impact of wholesaler second degree price discrimination by comparing model implied welfare under the observed price schedule to the counterfactual single price.

I find that banning quantity discounts decreases wholesaler profits as expected, increases retailer profits on average, and decreases consumer surplus. All three components are of similar magnitude. Thus, total welfare is reduced by removing quantity discounts. Welfare losses are larger for products with more heterogeneous downstream demand.

For wholesalers and retailers, the welfare consequences of banning price discrimination are similar to those seen in final goods contexts. Wholesalers always see a profit loss from being unable to price discriminate (this is a general proposition, see (Tirole 1988) §3). The effect on retailers depends on their size. Retailers who benefited from quantity discounts see input costs increase and profit losses, while the remaining retailers see the opposite. Effect magnitudes generally increase as the distribution of retailers becomes more heterogeneous. These conclusions are in line with those from the theoretical literature on the effects of price discrimination in a final goods context.

The impact on consumers comes through changes in retail prices. While many retailers see input cost declines after banning second degree price discrimination, the largest retailers see input cost increases and thus increase their prices. This raises the average retail price paid by consumers. The change in retailer prices is crucial to understanding the total welfare impacts of removing the upstream quantity discounts.

This paper is related to a long theoretical literature evaluating the impacts of price discrimination. Much of the theoretical work has been in final goods contexts, but the intuition provided in those settings can be helpful. This theoretical literature also serves as the basis for the structural model I develop. (Spence 1977) provides an early theoretical consideration of nonlinear pricing in a final goods market and derives predictions similar to the effects I find between wholesalers and retailers. Price discrimination increases profits for the seller (the wholesaler in the current context) and low value buyers see higher prices and lower consumption while high value buyers see discounts and increased consumption. (Katz 1983) examines nonlinear pricing in a final goods market and finds that the profit maximizing nonlinear price schedule distorts market consumption downward relative to the welfare maximizing schedule. (Maskin and Riley 1984) provide a comprehensive theoretical treatment of nonlinear pricing in a monopoly final goods context. They develop important technical conditions such as the single crossing property to ensure the presence of nonlinear pricing and derive theoretical results such as consumption being undistorted except for those consumers with the maximum marginal valuation (i.e. no distortion at the top). (Mussa and Rosen 1978) develop a similar framework for examining nonlinear pricing in quality and derive analogous predictions. (Wilson 1993) provides a comprehensive discussion of nonlinear pricing theory.

While the intuition and technical developments from a final goods context might be helpful, the present paper is concerned with an intermediate goods market. In such settings, the theoretical predictions are less clear. (Katz 1987) examines the impact of third degree price discrimination in an intermediate goods market in the context of the Robinson Patman Act and finds that allowing discrimination leads to higher input prices for all firms, lowering overall welfare. (O'Brien and Shaffer 1994) examine a similar context allowing firms to bargain over input prices and finds the opposite effect; that wholesalers being able to set different prices across firms can lower input prices for all firms and thus increases welfare. (Inderst and Shaffer 2009) present another theoretical discussion of the impacts of third degree price discrimination with two part tariffs in an intermediate goods market. They conclude that allowing the upstream firm to price discrimination increases welfare when they are allowed to offer simple two part tariffs. The current paper differs from this previous theoretical literature in that it explicitly considers second degree price discrimination in an intermediate goods industry and seeks to measure the welfare consequences empirically.

This paper is also the first to directly assess the impacts of wholesale quantity discounts on welfare. Previous empirical work on nonlinear pricing has been focused in final goods markets and examined the impact of other forms of nonlinear pricing. (McManus 2007) evaluates the impact of nonlinear pricing on product characteristics and finds nonlinear pricing distorts consumers away from their preferred product characteristics. (Nevo, Turner and Williams 2016) considers the welfare impacts of usage based pricing (another form of nonlinear pricing that offers a menu of free allocation, speed, and unit price) in the residential broadband market. They find that consumer surplus declines under usage based pricing, though total welfare increases. (Crawford and Yurukoglu 2012) examine the impact of unbundling television channels and find declines in consumer surplus but increases in overall surplus. Though not the main focus of the paper, (Villas-Boas 2009) infers the presence of price discrimination from differences in markups and examines the impact of a uniform markup. She finds that overall welfare increases when price discrimination is eliminated. (Grennan 2013) examines price discrimination in a bargaining context and finds that uniform pricing may decrease overall welfare, though it depends on the nature of the bargaining. These papers highlight that the empirical effects of banning price discrimination likely depend on market-specific factors as well as the nature of the price discrimination. This further emphasizes the point that modeling the multi-part tariffs seen in the data directly is important for any empirical conclusions.

Finally, this paper contributes to the literature studying a variety of economic issues in alcohol markets. (Conlon and Rao 2015) evaluate the effect of Connecticut's post and hold law on retail prices and consumption. (Cooper and Wright 2012) and (Saffer and Gehrsitz 2016) compare the effects of post and hold regulations across multiple states. (Miravete, Seim and Thurk 2017) examine the relationship between tax rates and tax revenue in the regulated Pennsylvania liquor market. (Aguirregabiria, Ershov and Suzuki 2016) consider partial deregulation and entry of private retailers in the Ontario wine market. (Seim and Waldfogel 2013) consider the impact of retail liquor store entry in Pennsylvania. (Ashenfelter, Hosken and Weinberg 2015) find evidence of efficiency gains in the national beer market form the merger of Miller and Coors. (Asker 2016) evaluates the effect of exclusive distribution arrangements in the Chicago beer market. This paper differs from this previous work on alcohol markets in that it considers the impact of nonlinear pricing at the wholesaler (distributor) level.

This paper is organized as follows. Section 2 gives relevant industry background information. Section 3 describes the data used in estimation. Section 4 describes the structural model in detail. Section 5 covers estimation and recovery of the supply side parameters. Section 6 presents estimation results and recovered structural parameter values. Section 7 considers the counter factual simulation of banning second degree price discrimination in the wholesale segment. Section 8 concludes.

1.2 Industry Background

In order to evaluate the welfare effect of prohibiting upstream second degree price discrimination, this paper considers wholesale discounts in the New York State liquor market. The primary feature of the liquor industry is that it is divided into three distinct tiers: distillers, wholesalers (or distributors), and retailers (i.e. package stores). The three tier structure is common across states and is not unique to New York. However, in New York, all three tiers are private industry (as compared with so-called control states like Pennsylvania or New Hampshire). In general, agents cannot belong to multiple tiers.[6] The practical implication of this regulation is that retailers and wholesalers can be treated as separate agents in the modeling process with no threat of merging.

Another relevant aspect of the liquor market in New York State is the lack of chain liquor stores as regulation prohibits an individual from owning more than one retail liquor license. This effectively prohibits the formation of chain retail liquor stores.[7] Additionally, liquor is unavailable in supermarkets. Therefore, consumers must go to a dedicated retail store to purchase liquor. These two effects together limit the negotiating power of any particular retail store with wholesalers. The implication for the model is that retailers will treat wholesale prices as given (as opposed to engaging in a bargaining process).

New York State also prohibits co-op buying at the retail level. That is, retail stores cannot band together to purchase a large amount of liquor and then divide it among themselves. This prevents retailers from coordinating to take advantage of quantity discounts.

[6] New York State Alcohol Beverage Control Law §101

[7] There appear to be small "chains" in practice. These seem to consist of separate licensees who engage in co-branding at the retail store level. The difficulty of coordinating decisions across a large number of individuals keeps the size of these limited.

Therefore, in the model, each retailer will be treated as making separate purchase decisions.

Finally, New York State has a "post and hold" law. Wholesalers are required to report their wholesale price schedules to the regulator on a monthly basis. Wholesalers are then required to adhere to these price schedules for the following month. The effects of post and hold regulation in Connecticut have been investigated by (Conlon and Rao 2015). Their setting is different in that Connecticut does not allow quantity discounts while New York does.[8] The post and hold regulation in New York is relevant because the "post" requirement means that the data is available to the relevant agents (i.e. retailers and wholesalers) and the researcher. The "hold" requirement means that wholesale prices are fixed for a given month. From the point of view of the retailer, this means that input costs are (relatively) fixed over a month. The fact that wholesalers are bound to these price schedules for a month also precludes the possibility of bargaining over wholesale prices.

Conlon and Rao argue that the post and hold system serves as a coordination mechanism for wholesalers. This stems from the fact that in Connecticut wholesalers are given a grace period after the initial posting to adjust their prices downward to match any competitor's prices.[9] New York's post and hold law has a similar provision. Conlon and Rao go on to argue that the ability to see and then match competitor's prices results in the wholesalers coordinating to the monopoly wholesale price. Similar intuition should apply here, though the pricing space is more complicated. New York wholesalers are able to post a price schedule, see their competitors' schedules, and then revise. The practical result is

[8]See Connecticut Agencies Regulations 30-6-A29(a) and New York State Alcohol Beverage Control Law §101.

[9]Conlon and Rao have the advantage of seeing both the initial and revised schedules. My data includes only the final schedules.

that the modeling will abstract from competition between wholesalers.[10]

1.3 Data

There are two primary classes of data that are used in the empirical part of this paper. The first is the wholesaler data which consists of the posted price schedules. The second is the retail demand data, which provides retail prices and purchase data.

1.3.1 Wholesale Data

If the goal is to understand the welfare impact of banning wholesale second degree price discrimination, it is important to understand the features of the quantity discounts that are offered. I collected a rich dataset of wholesale price schedules covering 2009-2013. The wholesale data originates from the New York State Liquor Authority (NYSLA), the industry regulator. The information was gathered through a freedom of information request. A non-machine readable version of the data is available on the NYSLA website.[11]

The descriptive statistics presented below serve to highlight that these price schedules look similar to those predicted by the theory of second degree price discrimination. (Wilson 1993) presents a framework for multi-part tariffs which look similar to those seen in the data. Further, the nature of the price discrimination is not related to observable characteristics of the products (except perhaps the prices which are likely related to downstream demand for those products). Wholesalers appear not to follow a rule of thumb or offer the same price

[10] Considering competition between wholesalers offering nonlinear prices is a difficult open problem. (Rochet and Stole 2002) and (Armstrong and Vickers 2001) examine oligopoly price setting behavior with nonlinear pricing and conclude that such problems are only tractable under strong simplifying assumptions.

[11] https://www.sla.ny.gov

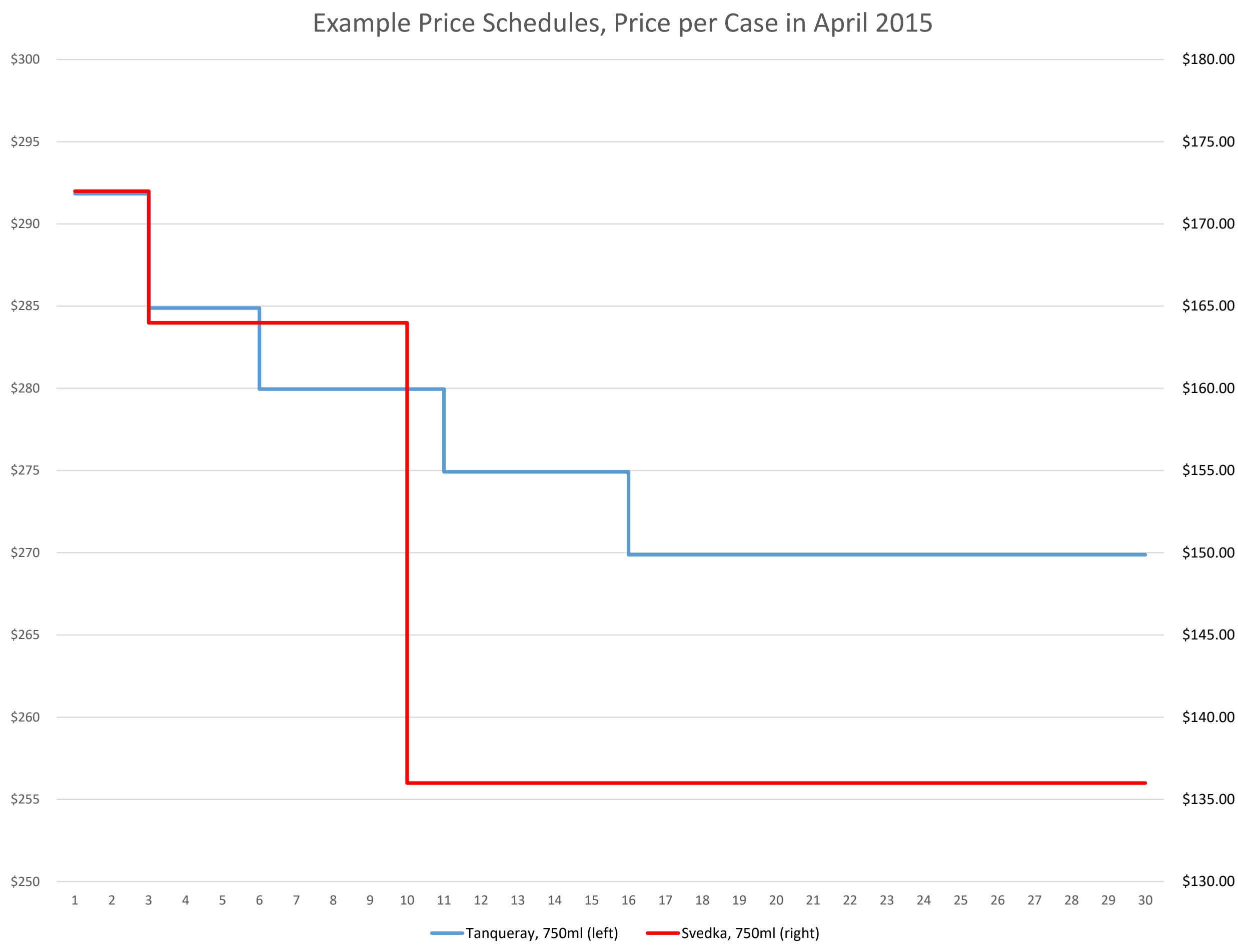
schedule for many different products.

Figure 1.1 presents examples of two price schedules. These examples are typical of the price schedules offered and suggest that wholesalers are trying to capitalize on retailer heterogeneity. The overall shape of the price schedules in Figure 1.1 are block declining tariffs, containing multiple price segments. Further, there are a finite number of segments in each tariff, suggesting that wholesalers are unable to fully exploit retailer heterogeneity. The above price schedule for Tanqueray has five segments (i.e. is a six part tariff) while the schedule for Svedka has three segments (i.e. is a four part tariff).[12] Also note that the quantity cutoffs differ across the products. Finally, the Svedka schedule exhibits larger changes between the prices offered while the Tanqueray schedule is more gradual in its changes.

Each brand and size combination has a price schedule reported monthly. For example, there are separate price schedules for Tanqueray in 1L and 1.75L bottles in addition to the price schedule presented in Figure 1.1. Price schedules are defined at approximately the same level as a UPC, which is the level at which the retail sales data is available. This enables the linking of particular retail demand data to a particular price schedule. While the price schedule and UPC definitions are essentially the same, there is no way to link the records directly. The price schedule information contains no UPC identifier requiring the two datasets to be hand matched. Defining price schedules at this level generates an enormous number of price schedules each month, another unique aspect of this dataset. Table 1.1

[12]By definition, an N part tariff consists of $N-1$ price options. Throughout the rest of the discussion the terms part and option will be used interchangeably unless the distinction is important.

Figure 1.1: Comparison of two price schedules.



The price per case for Tanqueray 750ml is on the left. The price per case for Svedka 750ml is on the right.

presents the total number of price schedules posted monthly over the entire sample period.[13] There are roughly 2,500 schedules posted monthly and the number posted is relatively stable.

Table 1.1: Number of price schedules posted per month.

Year	Month												Total
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
2009	1,612	1,377	2,463	2,461	2,608	2,578	2,589	2,521	2,577	2,587	2,608	2,594	28,575
2010	2,508	2,589	2,588	2,558	2,614	2,675	2,559	2,595	2,595	2,715	2,695	2,728	31,419
2011	2,611	2,719	2,891	2,802	2,859	2,856	2,238	2,208	2,287	2,293	2,345	2,373	30,482
2012	2,316	2,290	2,374	2,372	2,437	2,319	2,549	2,580	2,580	2,702	2,722	2,686	29,927
2013	2,647	2,707	2,607	2,731	2,679	2,723	2,833	2,754	2,762	2,916	2,950	2,958	33,267

The following tables present descriptions of the various segments of the price schedules. The data is broken down by the total number of segments available (i.e. by 2-part, 3-part, 4-part, etc. tariffs). Table 1.2 presents the mean, standard deviation, and median of each price option. Table 1.3 presents the same information on the cutoff quantities. These tables consider all price schedules denoted in cases from 2009-2013. The conclusion to draw from these tables is that there is substantial variation within a particular segment as well as across segments. Further, prices of a particular segment depend on the total number of segments offered. On the quantity side, while the medians appear similar, there is still important variation around that median.

In addition to variation in the prices across segments and by total number of segments, prices of particular segments are related to observable product characteristics. Table 1.4 presents the results of a regression relating prices to observable product characteristics

[13] This table and all following tables are for price schedules denoted in cases. These are typically more popular products (in the sense of brand recognition). The other option is to denote the price schedule in bottles. These typically are more expensive scotches or specialty liquors.

Table 1.2: Mean, standard deviation (in parentheses), and median price for each option by total number of options

Total Options	Schedule Option 1	2	3	4	5	6	7	8
1	251.56							
	(114.21)							
Obs. = 98	240.00							
2	175.93	153.50						
	(197.83)	(171.92)						
Obs. = 42,278	123.24	105.50						
3	162.30	148.87	138.68					
	(139.72)	(131.69)	(122.17)					
Obs. = 52,594	132.62	121.00	112.05					
4	175.72	163.02	154.10	142.96				
	(144.70)	(137.62)	(132.55)	(125.78)				
Obs. = 33,552	147.00	133.50	125.89	116.67				
5	193.92	182.44	174.07	165.05	154.57			
	(99.38)	(93.59)	(89.64)	(85.70)	(82.47)			
Obs. = 16,644	168.16	159.08	152.04	143.45	132.84			
6	203.05	192.34	184.55	176.54	167.59	157.46		
	(111.38)	(107.23)	(102.86)	(98.85)	(95.41)	(93.61)		
Obs. = 6,376	181.00	173.92	165.21	156.02	144.00	135.00		
7	201.28	189.98	183.35	177.00	169.64	162.03	154.72	
	(91.88)	(86.60)	(83.99)	(81.38)	(79.93)	(76.88)	(75.52)	
Obs. = 1,647	174.95	164.50	159.00	151.35	140.00	134.96	130.65	
8	158.76	148.73	142.66	136.54	128.87	118.41	96.19	71.03
	(39.92)	(38.71)	(37.03)	(36.10)	(38.94)	(46.15)	(104.47)	(173.20)
Obs. = 367	147.60	140.22	133.50	129.99	120.99	114.99	110.00	106.50

Table 1.3: Mean, standard deviation (in parentheses), and median cutoff quantity for each option by total number of options.

Total Options	Schedule Option 2	3	4	5	6	7	8
2	6.08						
	(22.49)						
N = 42,278	2.00						
3	4.86	8.31					
	(19.59)	(19.97)					
Obs. = 52,594	2.00	5.00					
4	3.98	6.68	14.81				
	(15.93)	(15.70)	(24.74)				
Obs. = 33,552	2.00	5.00	10.00				
5	5.11	7.81	13.73	28.17			
	(22.52)	(22.08)	(23.05)	(30.37)			
Obs. = 16,644	2.00	5.00	10.00	25.00			
6	4.13	7.12	13.17	24.58	50.51		
	(18.35)	(18.05)	(25.60)	(20.55)	(42.95)		
Obs. = 6,376	2.00	5.00	10.00	20.00	35.00		
7	2.60	5.25	10.53	20.30	37.60	72.05	
	(3.14)	(4.39)	(8.39)	(10.93)	(23.96)	(60.77)	
Obs. = 1,647	2.00	5.00	10.00	15.00	29.00	50.00	
8	3.06	5.43	10.32	20.20	33.71	64.96	88.70
	(0.96)	(1.31)	(2.61)	(12.73)	(18.72)	(70.58)	(52.31)
Obs. = 367	3.00	5.00	10.00	15.00	25.00	50.00	100.00

Note: Cutoff quantities are the minimum purchase required to get the price associated with that segment. The first cutoff quantity is always zero by definition as there is no minimum purchase quantity given.

controlling for which segments offers that price and total segments in the price schedule. Bottle size and spirit type are important determinants of price while proof is not. There also appears to be little to no seasonality in prices. The supply side parameters will be recovered product by product to control for these factors.

Table 1.4: Effect of Product Characteristics on Schedule Prices

Variable	Coef.	(SE)	p-value
1L	43.32	(0.52)	0.00
375ML	15.58	(0.65)	0.00
750ML	47.73	(0.49)	0.00
Proof	0.00	(0.01)	0.70
Gin	-48.96	(0.90)	0.00
Liqueur	-43.40	(0.80)	0.00
Rum	-51.77	(0.76)	0.00
Scotch/Whiskey/Bourbon	8.55	(0.68)	0.00
Tequila	-33.43	(1.03)	0.00
Vodka	-57.59	(0.69)	0.00
Constant	156.09	(1.44)	0.00
Controls for Option		Y	
Controls for Total Options		Y	
Controls for Month Effects		Y	

This table shows the results of a regression of price schedule prices on product characteristics. All prices regardless of segment are included. The regression controls for the segment of the particular price as well as the total segments offered.

1.3.2 Retailer Data

The strategy for understanding the welfare impacts of the quantity discounts offered by the wholesalers relies on developing a model to evaluate welfare with and without the quantity discounts. To capitalize on the unique nature of the price schedule data, the focus of the modeling and analysis will be on the wholesaler nonlinear pricing problem. Consumer demand is an input to that model. In contrast to the quality and extensive nature of the price schedule data, the retail sales data available to estimate consume demand is more limited. The goal is to find a reasonable representation of consumer demand that captures the relevant features given the limitations of the data.

The source of the retail data is the Nielsen Homescan Panel. The Nielsen Homescan Panel (hereafter the NHP) is a nationally representative sample of consumers. There are roughly 3,000 panelists annually in New York State, with approximately 800 consumers in New York City. Consumers are required to track all their purchases while they are in the panel. This consumption journal provides detailed information on every purchase a consumer makes.[14] Capturing all purchases by consumers is important as liquor is not sold in supermarkets in New York, excluding the product category from traditional scanner data sets.

Given the highly differentiated nature of liquor brands, the relatively small purchase amounts reported each month suggest that it may be unlikely to see small market-share products in a given month. Table 1.5 presents the number of purchases reported by panelists each month. Panel members purchase between 7,300 and 8,500 bottles annually. There appears to be some variation in total purchases across months with December and the summer months representing substantial increases over surrounding months.

With a relatively small number of purchases made each month, demand estimation

[14] Generally this is confined to goods with UPCs. There is marginal information on non-UPC goods such as produce.

will focus on the 250 most popular products among the most common sizes of 375ML, 750ML, 1L, and 1.75L bottles.[15] The top 250 products represents approximately 65% of bottles sold over the 2009-2014 time period among these sizes. Table 1.6 shows the breakdown of the top 250 products by spirit type.

Table 1.5: Liquor Bottle Purchases by Month

Year	Month												Total
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
2009	608	561	636	574	692	607	731	680	644	554	706	1,079	8,072
2010	547	502	543	506	662	616	686	647	540	545	634	965	7,393
2011	591	453	474	542	597	707	714	625	496	561	679	1,049	7,488
2012	630	540	618	619	788	886	892	813	573	527	566	1,014	8,466
2013	478	495	665	567	744	780	797	727	493	522	620	861	7,749
2014	509	507	598	601	710	773	901	750	550	524	610	738	7,771
Total	3,363	3,058	3,534	3,409	4,193	4,369	4,721	4,242	3,296	3,233	3,815	5,889	46,939

This subsample may vary in systematic ways from all products. Further, as noted above, only a subset of the most popular products will be able to be hand matched to corresponding price schedules. Table 1.7 presents summary statistics for each group. Note that the most popular products are similar in most dimensions except bottle size, with the largest bottle size being more common than among all products. The hand matched subsample appears similar to the top 250 products, though it appears to under represent the 375ML bottle size.

In addition to the purchase and product information, the NHP provides demographic data on the panel members. Information on race, age, education level, and employment are

[15]The other common bottle size excluded from this list is 50ML bottles. These bottles are extremely small and likely serve to target a different market which is cash constrained.

Table 1.6: Products by Spirit Type

Spirit Type	Number of Products
Gin	17
Vodka	63
Rum	32
Bourbon	31
Scotch	19
Whiskey	13
Tequila	7
Other	68

This table covers the 250 most popular products over 2009-2014. The "Other" liquor category corresponds to liqueurs, brandy, and other such products.

Table 1.7: Average Product Characteristics

Characteristic	All Products		Top 250		Matched	
	Mean	SD	Mean	SD	Mean	SD
Price	20.38	12.43	20.08	10.65	21.27	8.99
Proof	73.99	17.88	76.61	14.84	77.00	12.21
Imported	0.36	-	0.37	-	0.35	-
375ML	0.11	-	0.10	-	0.01	-
750ML	0.20	-	0.09	-	0.08	-
1L	0.24	-	0.22	-	0.24	-
1.75L	0.44	-	0.59	-	0.67	-

The "Matched" products represent 1,148 products hand matched to price schedules for 2011.

provided for all household members. Table 1.8 presents summary statistics regarding the income levels for all panel members in New York State and New York City in 2009-2013.[16] The relatively high fraction of high income households perhaps explains the higher fraction of larger bottle sizes. Demand estimation will control for consumer income.

Table 1.8: Consumer Income Distribution

Income (Dollars)	Statewide Percentage
2,500	0.34
6,500	2.36
9,000	1.21
11,000	2.63
13,000	3.91
17,500	2.02
22,500	4.99
27,500	9.84
32,500	2.76
37,500	2.83
42,500	4.45
47,500	2.49
55,000	11.73
65,000	4.38
85,000	24.14
150,000	19.89

1.4 Model

This section develops the structural model used to evaluate the welfare impacts of prohibiting second degree price discrimination in the wholesale market. The structural model has three main components: wholesaler behavior, retailer behavior, and consumer demand.

[16] Income is reported as a series of bands. The data bellow is created by taking the midpoint of each band.

Figure 1.2 presents the general structure of the model along with the model timing. Consumers have preferences over various liquor types which generate purchases of liquor. Consumer and retail stores are located around a limited geographic area (for example, New York State). The spatial distribution of consumers and retailers generates variation in local demand at a particular establishment and determines the "type" of a particular retailer. There is a single wholesaler who knows the distribution of this local demand variation but does not know the realization at any one location. The wholesaler engages in second degree price discrimination by offering nonlinear price schedules to maximize expected profits, taking advantage of this heterogeneity in local demand.[17] Each component will be discussed in turn.

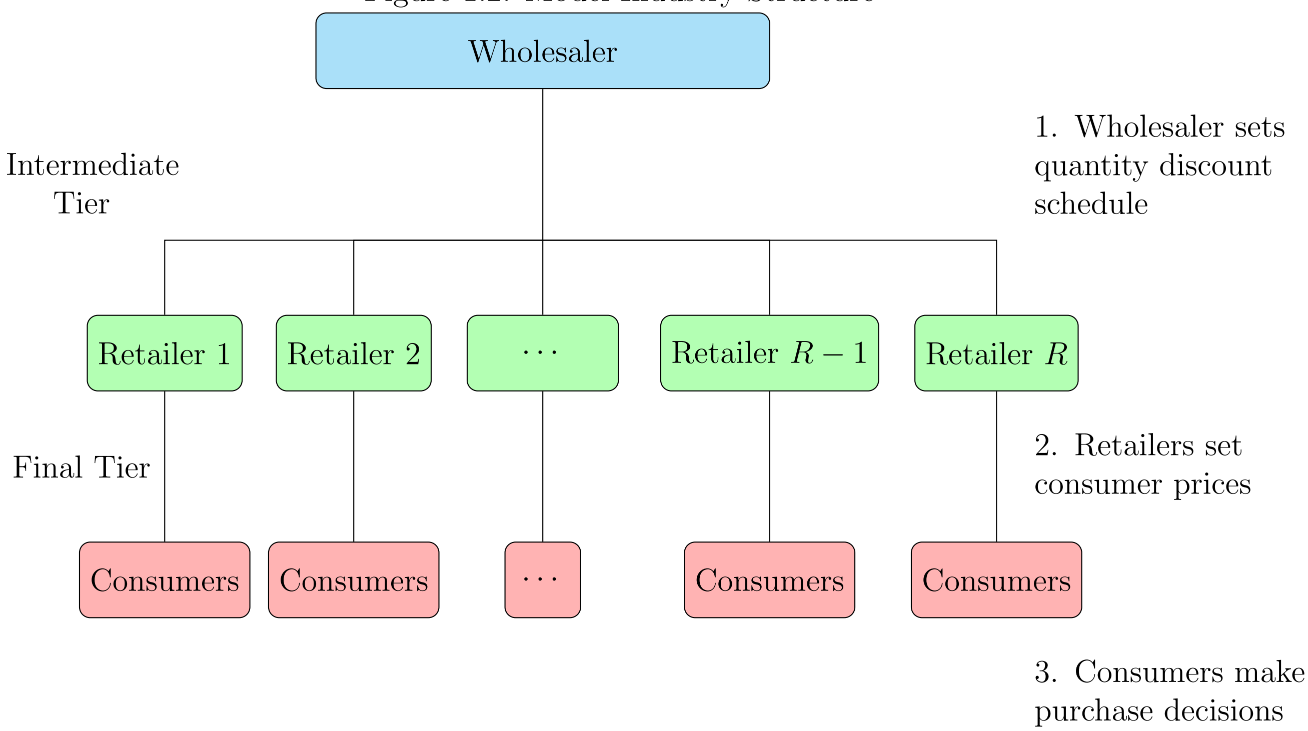
1.4.1 Fundamentals

Consider a monopolist wholesaler who has marginal cost c_j for each product j that faces a large number of downstream retailers who are each local monopolists. Retailers are heterogeneous and each retailer is characterized by a vector of types λ_{rj}, one for each product, each independently drawn from some product specific bounded distribution $F_j(\lambda_{rj})$. Without loss of generality, I will assume that $F_j(\lambda_{rj})$ is defined on $[0, 1]$. In each market, consumers generate some downward sloping market share function for each product $s_j(p)$.

The type λ_{rj} represents a local scaling of demand for a particular product at a particular retailer. For example, variation across retailers (but within a product j) might arise from geographic dispersion of consumers, variations in the income distribution across local markets, or variation in retail experience. The interpretation is that the type λ_{rj} is a

[17] Alternatively, the wholesaler knows the realization at any particular location but is prevented from engaging in perfect price discrimination.

Figure 1.2: Model Industry Structure



local demand level that the retailer knows better than the wholesaler. That is, retailers' know their own types while the wholesaler only knows the distribution $F_j(\lambda_{rj})$. Knowing only the distribution of types means that the wholesaler cannot engage in first degree price discrimination. The wholesaler therefore uses nonlinear pricing for each product to price discriminate over this distribution.

1.4.2 Retailer Problem

Retailers are assumed to engage in monopolistic competition. (Tirole 1988) §7.2 gives the distinguishing features of monopolistic competition as downward sloping firm demand and negligible strategic interaction between many firms. Data on retailer specific sales in the Nielsen Homescan Panel is limited. While the monopolistic competition assumption is extreme in the sense that it abstracts from competition, it does incorporate retailer heterogeneity and double marginalization which are critical to understanding the impact of quantity discounts in an intermediate goods industry. Further, the data to support a different assumption is not available. The monopolistic assumption will induce the model-predicted retail prices to be too low. The magnitude of this effect depends on the cross price elasticities of the different products.

The downward sloping demand assumption amounts to assuming that retailers have price setting power. One can assume that retailers are differentiated in some sense, certainly by location and perhaps by retail experience or staff knowledge. In the presence of non-negligible search or transportation costs, each retailer would have at least some price setting power. Analysis of liquor license data indicates that there are more than 3,000 liquor stores in the state with 1,412 liquor stores in the five counties composing New York City. Thus, the “many firms” clause is satisfied. Therefore, while monopolistic competition is a potentially limiting assumption, it is consistent with the structure of the market.

The model also abstracts from any multi-product price setting concerns. The assumption will be that retailers will maximize profit across products independently. Both (Conlon and Rao 2015) and (Miravete et al. 2017) find very small cross price elasticities across spirits. Therefore, the effect of this assumption should be relatively minor. In the discussion below, the product subscript is suppressed.

With the above assumptions about competition and price setting behavior, the retailer profit can now be described. Retailers face a price schedule set by the wholesaler. The price schedule is a multi-part tariff containing a finite number of segments and can be characterized by a series of two part tariffs consisting of a fixed fee and linear price within that segment. Choosing to be in a segment $k = 1, \ldots, N-1$ means that the retailer's total cost is $A_k + \rho_k q$, where q denotes the amount purchased by the retailer. Given that a retailer chooses to be in a particular segment k, their profit is given by:

$$\pi_r(p; \rho_k) = (p - \rho_k)s(p)\lambda_r M - A_k \,, \tag{1.1}$$

where the retail price is denoted by p, and M represents a general demand scaling parameter common across retailers. The presence of M allows λ_r to have support bounded on $[0, 1]$. Thus, the amount sold by a retailer is $s(p)\lambda_r M$. It should be noted that choosing a particular tariff segment is dual to purchasing the amount required to fall into that segment. The problem could be recast in terms of picking the optimal quantity given the tariff, but would lead to the same solution.

The profit maximizing price given for a retailer choosing segment k is characterized by:

$$(p^\star - \rho_k)s'(p^\star) + s(p^\star) = 0\,, \tag{1.2}$$

which implicitly defines a function $p^\star(\rho_k)$ giving the optimal retail price as a function of the wholesale price for that segment. Note that the optimal price does not depend on the type λ_r. Therefore, all retailers in a particular segment charge the same price. All else equal, if retailers have similar input costs one would expect them to charge similar prices. The relatively low number of retail prices that will be charged in equilibrium is supported by the retail data. Figure 1.3 shows the distribution of unique prices observed for a product in a month. Relatively few retail prices are observed in practice.

Though they charge the same price, profits would vary across retailers according to variation in λ_r. To see this, plug in the the optimal retail price function into the retailer profit shown in Equation 1.1. In equilibrium, the amount purchased by a retailer is $s(p^\star(\rho_k))\lambda_r M$ and the corresponding profit for being in that segment as:

$$\pi_r^\star(k) = (p^\star(\rho_k) - \rho_k)s(p^\star(\rho_k))\lambda_r M - A_k\,, \tag{1.3}$$

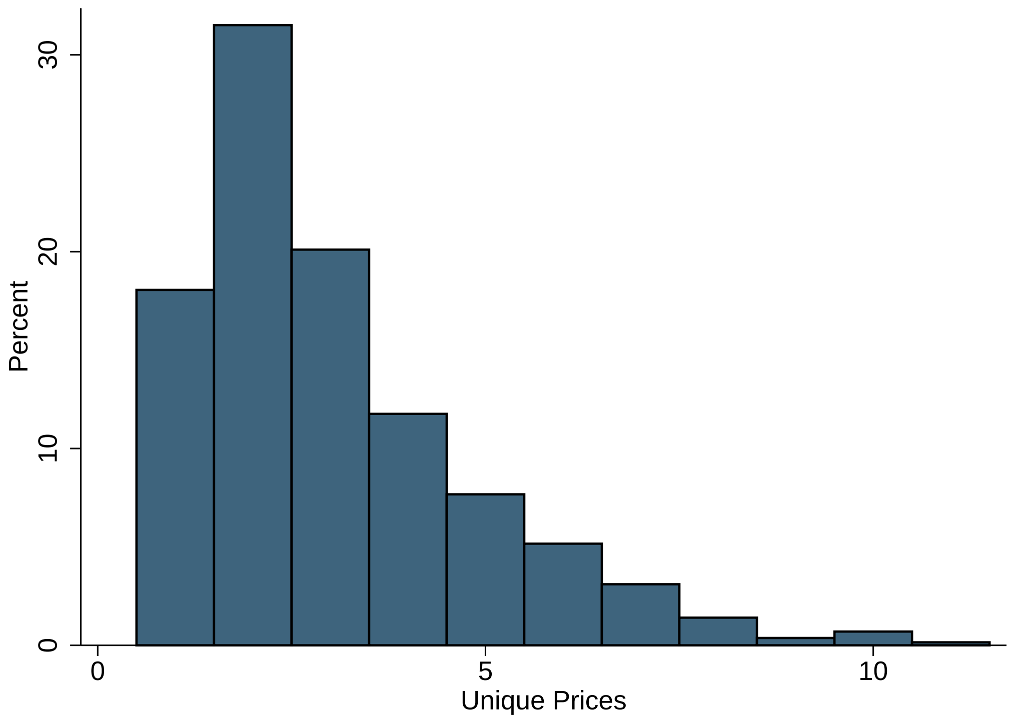
which clearly depends on λ_r.

1.4.3 Wholesaler Problem

The wholesaler sells a large number of products, each indexed by j. For consistency with the retail model, the wholesaler maximizes profit for each product independently, ignoring cross product effects. The product subscript is suppressed in the discussion below.

The wholesaler problem is to pick the profit maximizing price schedule with $N - 1$

Figure 1.3: Distribution of Unique Retail Prices by Product by Month



Limited to top 250 products observed more than once in a month.

prices given the distribution $F(\lambda_r)$, a constant marginal cost c, retailer behavior $p^\star(\rho_k)$, and consumer demand $s(p)$.[18] F and c are assumed to be product specific.

The wholesaler has several choice variables. Conditional on the number of segments in the price schedule, the wholesaler must choose the linear price within each segment ρ_k and the associated fixed fee A_k for each segment $k = 1, \ldots, N-1$. Within a segment, the wholesaler gets profit from a retailer of:

$$(\rho_k - c)s(p^\star(\rho_k))\lambda_r M + A_k \, . \tag{1.4}$$

Recall that the retailer type parameter λ_r represents local demand. Assuming the wholesaler only knows the distribution $F(\lambda_r)$ amounts to assuming that the wholesaler doesn't observe local demand but may have some belief over potential outcomes.

It is important that F have an increasing hazard rate. (Wilson 1993) §8 points out that that an increasing hazard rate is a sufficient condition for the optimal purchase size to be non-decreasing in type. This property is satisfied by many common distributions, such as the uniform distribution. I will assume that F is a particular family of Beta distributions which satisfy the increasing hazard rate property. Specifically, I assume:

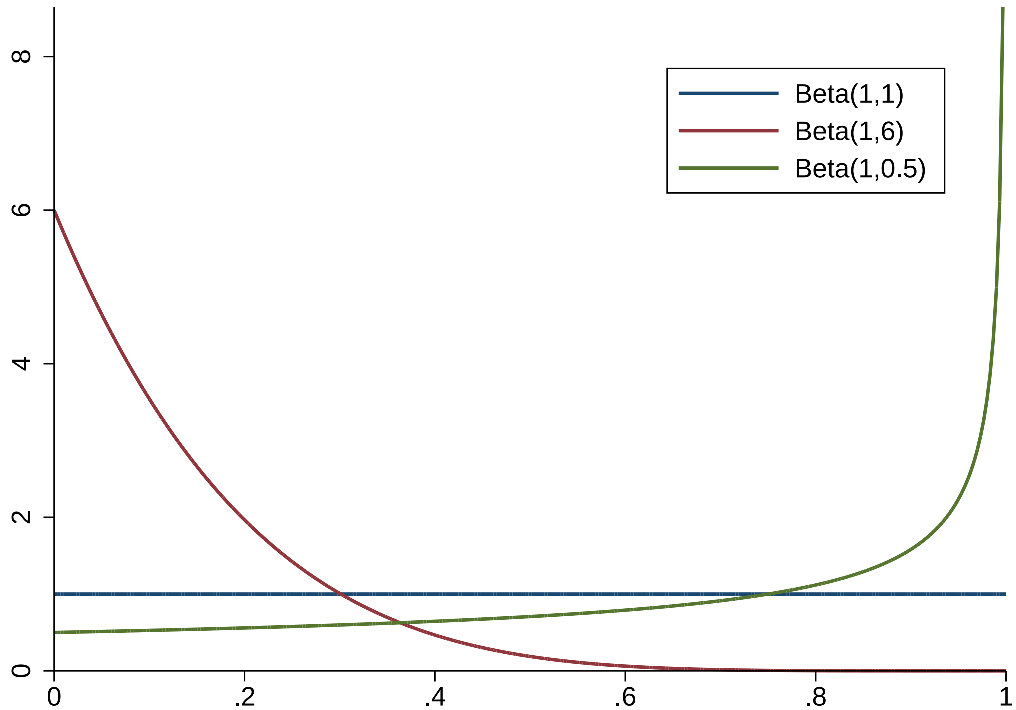
$$F(\lambda; b) = 1 - (1-\lambda)^b. \tag{1.5}$$

This is equivalent to assuming that $F \sim Beta(1, b)$. Recovering te first parameter in addition to b (while still ensuring the increasing hazard rate property) does not qualitatively change the results of the counterfactual simulations. Figure 1.4 provides three examples from this

[18] The model will treat the number of segments offered as exogenous. That is, the number of prices offered by the wholesaler will be assumed to be the number observed in the data.

family of distributions, the uniform ($b = 1$) and two others. As b increases, the distribution becomes less heterogeneous and shifts towards zero. Allowing $b < 1$ shift mass towards the high end of the distribution.

Figure 1.4: Comparison of Densities for Different Values of b



It is also important that λ_r enters multiplicatively in the retailer profit function. This assumption ensures that the demand functions for retailers (i.e. the intermediate demand) do not intersect. Retailers with a higher type will buy more at any given wholesale price. (Maskin and Riley 1984) note that this so-called single crossing property guarantees sorting of consumers in equilibrium. (Wilson 1993) §6 notes that sorting is crucial to ensure that the optimal nonlinear price schedule is monotonic. Combined with the increasing hazard rate property, the single crossing property ensures that the optimal nonlinear price schedule is one with quantity discounts.

Assuming that retailer demand is given by $\lambda_r s(p^\star(\rho))M$ ensures the single crossing property and that retailer demand can be ordered by type. Retailers purchasing from segment k are those retailers with types $\lambda_k \leq \lambda_r < \lambda_{k+1}$. These cutoff types define the quantities at which the price schedule will change. The quantities cutoffs defined by the type cutoffs are the model analogue of the quantity thresholds seen in Figure 1.1.

This allows us to define the profit a wholesaler expects from a segment:

$$\int_{\lambda_k}^{\lambda_{k+1}} \left[(\rho_k - c))s(p^\star(\rho_k))\lambda_r M + A_k\right] dF(\lambda_r)\,. \tag{1.6}$$

The profit from the whole schedule can thus be given by:

$$\pi_w = \sum_{k=1}^{N-1} \int_{\lambda_k}^{\lambda_{k+1}} \left[(\rho_k - c))s(p^\star(\rho_k))\lambda_r M + A_k\right] dF(\lambda_r)\,. \tag{1.7}$$

The wholesaler's choice variables are the prices ρ_k, fixed fees A_k, and cutoff types λ_k.

To ensure continuity of the price schedule, additional constraints must be imposed. Intuitively, the retailer with the cutoff type should be indifferent between their assigned segment k and the one below $k-1$ (otherwise it is not a cutoff type). That is, the profit for a retailer of the cutoff type should be the same for being in segment k or segment $k-1$. Formally, this gives the set of conditions:

$$(p^\star(\rho_k) - \rho_k)s(p^\star(\rho_k))\lambda_k M - A_k = (p^\star(\rho_{k-1}) - \rho_{k-1})s(p^\star(\rho_{k-1}))\lambda_k M - A_{k-1}\,. \tag{1.8}$$

These conditions can be used to eliminate the choice of fixed fees from the wholesaler's program. The profit function can be reformulated entirely in terms prices and cutoff types.

After substituting in these conditions defining the fixed fees, the wholesaler profit function is given by:

$$\begin{aligned}\pi_w = &\sum_{k=1}^{N-1} \int_{\lambda_k}^{\lambda_{k+1}} \left[(\rho_k - c))s(p^\star(\rho_k))\lambda_r M\right] dF(\lambda_r) \\ &+ (1 - F(\lambda_k))\lambda_k M((p^\star(\rho_k) - \rho_k)s(p^\star(\rho_k)) - (p^\star(\rho_{k-1}) - \rho_{k-1})s(p^\star(\rho_{k-1})))\,.\end{aligned} \tag{1.9}$$

The first term represents profits from unit sales. The second term represents the profit from fixed fees. To fix ideas further, it is instructive to consider a simple two part tariff (i.e. $N = 2$). In this case, all retailers are subject to the same tariff given by A_1 and ρ_1. The wholesaler's expected profit is given by:[19]

$$\pi_w = \int_{\lambda_1}^{1} \left[(\rho_1 - c)s(p^\star(\rho_1))\lambda_r M\right] dF(\lambda) + (1 - F(\lambda_1))\lambda_1 M((p^\star(\rho_1) - \rho_1)s(p^\star(\rho_1)))\,. \tag{1.10}$$

The integral term represents the profit gained from selling units to all the types participating in the market ($\lambda_r > \lambda_1$). The second term represents profit gained from the fixed fee associated with that price, which in this case is the profit of the lowest participating type. price. The first order conditions with respect to the price and cutoff types characterize the (unique) optimal schedule. For this simple two part tariff, the optimality conditions are:

$$\begin{aligned}&\left[(\rho_1 - c)s'(p^\star(\rho_1))\frac{\partial p^\star}{\partial \rho_1} + s(p^\star(\rho_1))\right] \int_{\lambda_1}^{1} \lambda_r dF(\lambda_r) \\ &\qquad + \lambda_1(1 - F(\lambda_1)) \left[(p^\star(\rho_1) - \rho_1)s'(p^\star(\rho_1))\frac{\partial p^\star}{\partial \rho_1} + s(p^\star(\rho_1))(\frac{\partial p^\star}{\partial \rho_1} - 1)\right] = 0\,,\end{aligned} \tag{1.11}$$

$$-\lambda_1 f(\lambda_1)(\rho_1 - c)s(p^\star(\rho_1)) + [1 - F(\lambda_1) - \lambda_1 f(\lambda_1)]\left[(p^\star(\rho_1) - \rho_1)s(p^\star(\rho_1))\right] = 0\,. \tag{1.12}$$

[19]The tariff "option" with $k = 0$ is considered non-participation. Not participating thus generates zero profit for the retailer.

For a multi-part tariff, there will be a total of $2(N-1)$ first order conditions. These first order conditions characterize the optimal prices and cutoffs for each segment. In general, the first order conditions describing the prices are given by

$$\begin{aligned}\left[(\rho_k - c)s'(p^\star(\rho_k))\frac{\partial p^\star}{\partial \rho_k} + s(p^\star(\rho_k))\right] \int_{\lambda_k}^{\lambda_{k+1}} \lambda_r dF(\lambda_r) \\ + \left[(p^\star(\rho_k) - \rho_k)s'(p^\star(\rho_k))\frac{\partial p^\star}{\partial \rho_k} + s(p^\star(\rho_k))(\frac{\partial p^\star}{\partial \rho_k} - 1)\right] \\ \times \left[\lambda_k(1 - F(\lambda_k)) - \lambda_{k+1}(1 - F(\lambda_{k+1}))\right] = 0\,, \end{aligned} \quad (1.13)$$

and the conditions for the optimal cutoffs are given by

$$\begin{aligned} &-(\rho_k - c)s(p^\star(\rho_k))\lambda_k f(\lambda_k) + (\rho_{k-1} - c)s(p^\star(\rho_{k-1}))\lambda_k f(\lambda_k) \\ &\quad + [1 - F(\lambda_k) - \lambda_k f(\lambda_k)]\left[(p^\star(\rho_k) - \rho_k)s(p^\star(\rho_k)) - (p^\star(\rho_{k-1}) - \rho_{k-1})s(p^\star(\rho_{k-1}))\right] = 0\,. \end{aligned} \quad (1.14)$$

1.4.4 Consumer Demand

Up until this point, I have assumed that consumers generate some market share function $s_j(p)$ that describes demand for a particular product. I now describe the consumer behavior that generates this demand function.

I follow most of the empirical industrial organization literature and derive consumer demand from a discrete choice model of individual level product choice. The liquor market consists of a large number of products when considered at the UPC level. To model consumer choice across all possible products would be a daunting task indeed. As described in 1.3.2, the empirical specification considers the most popular 250 liquor UPCs. I also include an outside option (denoted as product "0") which indicates choosing something other than one of the 250 most popular products or no liquor at all.

Each consumer i chooses a product $j \in \{0, 1, \ldots, 250\}$ in month t. They get utility U_{ijt} defined as:

$$U_{ijt} = \alpha_1 p_{jt} + \alpha_2 p_{jt} \ln(\frac{income_i}{10000}) + x_j \beta + \epsilon_{ijt} \qquad \forall j \in \{1, \ldots, 250\} \tag{1.15}$$

$$U_{i0t} = \epsilon_{i0t} \qquad j = 0 \tag{1.16}$$

where $income_i$ denotes the income of consumer i. The $x_j \beta$ term represents the impact of product characteristics on consumer utility. During estimation, these will be controlled for with a series of product dummies. This also captures any time invariant unobserved product attributes relevant to consumers. Unobservable (to the econometrician) components of utility are given by ϵ_{ijt}.

To complete the demand specification, an assumption is required on the distribution of the unobservable components of utility. I assume that the ϵ_{ijt} follow a Type-I extreme value distribution such that the choice probabilities of the above model are given by a logit model.

As described in Section 1.3.2, the Nielsen Homescan Panel provides individual level purchase decisions. The parameters of the model can thus be estimated via maximum likelihood. Identification of model parameters will come from differences in purchase probabilities across consumers rather than across markets. (Shum 2004) provides an example of estimating a similar demand system using individual level data. See (Train 2009) for a discussion of the properties and estimation of these and related models.

One limitation of the above specification is the inability to control for product-month specific unobserved effects. For example, an advertising campaign for Tanqueray in April is not captured by the observables in the model. Importantly, such unobserved effects might be relevant for pricing decisions by firms. That is, prices may be correlated with these product-time varying unobserved effects. This correlation between prices and unobservable characteristics creates an endogeneity problem that must be dealt with in estimation.

The typical approach to addressing endogeneiety in discrete choice models is to implement some version of the estimation approach outlined in (Berry, Levinsohn and Pakes 1995), (Berry, Levinsohn and Pakes 2004), or other related approaches. Critical to these types of approaches is the assumption that market share data is observed with no error. If enough individual level data is available, individual level choices may be aggregated to get a measure of product market share. However, the Nielsen Homescan Panel does not provide enough individual purchases of each product in a particular month such that aggregating the individual choices represent an accurate measure of the product market share. Therefore, such aggregate level approaches are not appropriate given the data available.

(Petrin and Train 2010) propose an alternative approach that relies on control function methods. Their approach is valid in individual level models that do not rely on observing aggregate market shares. They also demonstrate that their approach provides similar results to the aggregate level methods described above when the necessary assumptions about market share data are valid. The control function method presents a plausible way to control for endogeneity in the model specified above.

1.4.5 Computational Considerations

Before addressing extensions to this model, it is worthwhile to discuss difficulties that arise in this setting. The function $p^{\star}(\rho)$ has no closed form solution and must be found numerically because consumer demand is derived from a logit formulation. This adds a computational burden to evaluating the first order conditions given above. The function $p^\star(\rho)$ enters both directly and in the form $s(p^\star(\rho))$. In general, if the consumer demand is nonlinear then the ρ_k term will enter nonlinearly in the first order conditions. This complicates the solution of the first order conditions for the optimal tariff. Finally, the optimality conditions obtained from the above profit function include an integral term with one of the choice variables in the bounds. Thus, unless the integral $\int \lambda dF(\lambda)$ has a closed form, a closed form solution for the optimal price schedule is not possible.

The difficulties pointed out above are not prohibitive to using this framework. They present computational rather than theoretical challenges. Solving the above framework with functional forms useful for empirical work is possible, but necessitates numerical solutions. Further, tight tolerances and careful techniques must be employed to ensure an optimum is found. Appendix B discusses these issues further and details the computational approach used in practice.

1.4.6 Extensions

The setting presented above covers most of the features relevant to measuring the welfare consequences of preventing the wholesaler from engaging in price discrimination. There are three main extensions to the setting presented above: including cross product effects, wholesaler competition, and retailer competition.

The first extension is to consider pricing behavior that includes cross product effects. This affects the above model in two ways. First, retailer price setting would depend on the cross-price elasticities of the various products. The retailer prices would also depend on the marginal cost of each product (i.e which segment of the price schedule the retailer is on for every product they sold). Second, because the retailer price setting behavior $p^\star(\rho)$ depends on all the price schedules so will the solution for the optimal price schedule for each product. At this point, the model becomes a multidimensional screening model. While multidimensional screening has been studied theoretically, such an approach is not feasible for empirical work. (Wilson 1993) §14 covers such a model and provides only partial characterization. (Armstrong 1996) considers multidimensional nonlinear pricing and is able to characterize general results for subset of models satisfying some separability conditions. The focus of this paper is on the impact of nonlinear pricing. Including multidimensional screening complicates the problem without addressing the fundamental question.

The second extension is to consider competition between wholesalers who set nonlinear price schedules. (Rochet and Stole 2002), (Armstrong and Vickers 2001), and (Stole 1995) present theoretical evaluations of nonlinear pricing in an oligopoly context (primarily with single products, only two firms, and only with completely nonlinear schedules). In all cases, equilibria are difficult to characterize and strong assumptions such as symmetry and full market coverage are required. The nature of these theoretical models makes them ill-suited for empirical use. Further, these models largely focus on the single product case reflecting the difficult nature of multiproduct competition in nonlinear pricing.

Table 1.9 presents a measure of competition between wholesalers available from the price schedule data. The idea is to try and measure the frequency with which a particular product is offered by a given number of wholesalers. It is important to consider the timing aspect of this as well, as it appears that the products offered by a wholesaler change over time. The table presents the number of product-months for which the reported number of wholesalers offer the same product. A product is defined as a brand-size combination (i.e. Tanqueray 750ML). For example, if products were only offered by one wholesaler every month, then all observations would be in the top row. Alternatively, if every month five wholesalers all offered the same products, then all observations would be in the bottom row. As can be seen, more than half of product-months are offered by just one wholesaler, while almost all are offered by two or fewer wholesalers. That is, the vast majority of the time a product is offered by only one or two wholesalers state-wide.

Table 1.9: Distribution of Number of Wholesalers Selling a Product

# of Wholesalers	Product-Months	Percentage
1	182,704	57.52
2	134,869	42.46
3	64	0.02
4	14	0.00
5	7	0.00

The final extension to consider is competition between retailers. The model outlined above assumes that the retail market is monopolistically competitive. Introducing competition between retailers raises a number of questions. From a consumer's point of view, retailers are offering essentially homogeneous products. However, retailers may be differentiated in their "retail experience" (e.g. staff knowledge, location, the pleasantness of the shopping environs, etc.). Modeling competition between retailers must account for two issues: consumers' preferences over retailers for identical goods, and how retailers best respond to one another. Ultimately, retail competition models reduce to local monopoly pricing on residual demand. The differences are in how that residual demand arises. The above model captures variation in the residual demand through the retailer type λ_r.

1.5 Recovering Parameters

The goal of this paper is to compare the welfare under a regulatory regime that allows quantity discounts versus a regime that allows only linear pricing. There is no exogenous policy change during the sample that would allow direct estimation of the effect of banning nonlinear pricing. Rather, the structural model presented above can be used to evaluate how different wholesale pricing behaviors impact welfare. The above model nests both cases of interest; it allows for both price schedules with several options, as observed in the data, and only linear prices (by setting $N = 2$ and constraining fixed fees to be zero).

I recover the parameters of the structural model (conditional on the demand specification) from the shape of the observed price schedules, schedule by schedule. While the consumer demand system is estimated across all products, the wholesaler marginal cost, complexity cost, and retailer distribution parameter is recovered separately for each price schedule. Intuitively, this is similar to inverting the model to recover the parameters. This allows the supply side parameters to vary freely across products and over time. I rely on inequalities implied by the profit maximization behavior of the wholesaler to recover the desired parameters. This approach exploits the rich nature of the price schedule data. With the estimated parameters, the welfare under each regulatory regime can be compared.

There are two sets of parameters to recover: the parameters of the consumer utility function and the supply side parameters. Consumer utility parameters are common across consumers and products. The wholesaler model introduces product specific parameters: the wholesaler marginal cost c, the type cutoffs $\lambda_1, \ldots, \lambda_{N-1}$, and the parameters of the distribution of types $F(\lambda)$ to be recovered.[20] The retail model contains no additional parameters to estimate. Product characteristics (including price), consumer demographics, and consumer purchase decisions are observed from the NHP data. The fixed fees and linear prices (A_k and ρ_k) can be recovered directly from the reported price schedules. Variation in consumer purchase decisions and product characteristics allows the parameters of the utility function to be estimated.

1.5.1 Demand Estimation

The utility model specified above implies a series of individual choice probabilities for each product. The NHP data consists of a series of individual purchases with linked demographic data and product characteristics. The parameters of the utility function can be estimated via standard maximum likelihood techniques. The estimation sample consists of all liquor purchases observed in the data as well as trips for which consumers purchased wine but not liquor. Wine is also not sold in supermarkets and so serves as a reasonable proxy for a consumer trip to a liquor store. Inclusion of these trips helps identify the value of the top 250 products relative to the outside option. As noted above, the demand estimation recovers parameters for price, a price-income interaction, and a series of product dummies for the most popular products. These product dummies are then be projected onto product characteristics to recover the effect of product characteristics on demand.

[20]To aid clarity, the product and time subscripts are suppressed.

1.5.2 Supply Side Parameters

As discussed in Section 1.4.3, a parametric assumption was made on the distribution of retailer types. The type distribution $F(\lambda_r)$ was assumed to follow a restricted form of the Beta distribution such that $F(x) = 1 - (1-x)^b$. This distribution is flexible and can take a number of shapes over the support $[0, 1]$. This introduces one distribution parameter to recover. Define the set of supply side parameters to be recover as $\theta_s = \{c, b, \lambda_1, \ldots, \lambda_{N-1}\}$.

The type cutoffs can be directly inferred from the indifference conditions of the retailer. As before, the cutoff type must be indifferent between the profit at segment k and segment $k-1$. The resulting indifference condition can be solved for the cutoff type. That is:

$$\lambda_k = \frac{A_k - A_{k-1}}{M\left[p^\star(\rho_k) - \rho_k)s(p^\star(\rho_k)) - (p^\star(\rho_{k-1}) - \rho_{k-1})s(p^\star(\rho_{k-1}))\right]}, \tag{1.17}$$

One must assume that $A_0 = 0$. It should also be noted that for all observable price schedules $A_1 = 0$ because there is no fixed fee just to see the offerings of the wholesaler.[21]

I employ a moment inequality approach to recover the remaining parameters (c, b). The following discussion follows the ideas of (Bajari, Benkard and Levin 2007) and (Pakes, Porter, Ho and Ishii 2015). The required assumption is that the observed price schedules are a profit maximizing choice by the wholesalers. Thus, at the true value of the parameters, the observed price schedule must be more profitable than any other price schedule. Denote a price schedule $\omega = \{\rho_1, \ldots, \rho_{N-1}, \lambda_1, \ldots, \lambda_{N-1}\}$ and the observed price schedule as ω^0. Then

[21] M is calibrated to be 2,000,000 for all price schedules. This value translates to every adult over 21 in New York State purchasing approximately 2 bottles of liquor annually - an amount close to the median value in the data.

the above intuition can be put formally as:

$$\pi^w(\omega^0, N; \theta_s^0) \geq \pi^w(\omega', N; \theta_s^0)\,, \tag{1.18}$$

for any other price schedule ω', observed number of segments N, and true parameter vector θ_s^0.

While the above inequality holds at the true value of the parameters, the set of parameters consistent with the above inequalities might be larger than a singleton. That set can be defined as:

$$\Theta_s = \left\{\theta_s : \pi^w(\omega^0, N; \theta_s) \geq \pi^w(\omega', N; \theta_s)\; \forall\, \omega'\right\}\,. \tag{1.19}$$

To find this set, a measure of violations of the inequality is minimized. For a particular deviation of the price schedule $\omega_i' \neq \omega^0$ we can define ν_i as the amount of violation of the above inequality. That is:

$$\nu_i(\theta_s) = \min\left\{0, \Pi^w(\omega^0, N; \theta_s) - \Pi^w(\omega_i', N; \theta_s)\right\}\,, \tag{1.20}$$

where $\nu_i(\theta_s)$ takes the value zero if θ_s is consistent with the above inequality and takes a negative value otherwise (i.e. if the deviation price schedule generates more profit at that guess of θ_s). This naturally leads to the following objective function to be minimized:

$$Q(\theta_s) = \sum_i \nu_i^2\,. \tag{1.21}$$

The zero set of the above function is identical to the set Θ_s.

The above approach relies on defining a set of deviation price schedules. To see how this is done, consider a three part tariff, consisting of two prices $(\rho_1, \rho_2) = (6.00, 2.00)$ and two type cutoffs $(\lambda_1, \lambda_2) = (0.20, 0.60)$ so $\omega^0 = (6.00, 2.00, 0.20, 0.60)$. Fix a deviation amount of $\delta = .1$. Then a series of deviation price schedules can be constructed by multiplying each component of the observed price schedule by $1 + \delta = 1.1$. For example, employing this process to each element of the price schedule in turn gives:

$$\omega_1' = (6.60, 2.00, 0.20, 0.60) \tag{1.22a}$$

$$\omega_2' = (6.00, 2.20, 0.20, 0.60) \tag{1.22b}$$

$$\omega_3' = (6.00, 2.00, 0.22, 0.60) \tag{1.22c}$$

$$\omega_4' = (6.00, 2.00, 0.20, 0.66) \tag{1.22d}$$

Another set of deviation price schedule can be considered by perturbing two parts of the price schedule (i.e. multiplying two elements of ω^0 by $1+\delta$). For example, $\omega_5' = (6.60, 2.00, 0.22, 0.60)$. Likewise for three and four elements.

An entire other set of perturbed price schedules can be generated by multiplying elements by $1 - \delta$ instead, scaling down each part of the price schedule. Finally, a mixture of scaling up and down (i.e. multiplying some elements by $1 + \delta$ and others by $1 - \delta$) can generate even more perturbations. An example of such a perturbation would be $\omega_6' = (5.40, 2.20, 0.22, 0.54)$.

By fixing a simple scaling factor δ a large number of price schedule perturbations can be generated. In general, this procedure can generate $3^{2(N-1)} - 1$ deviation price schedules for a fixed δ. For the purposes of the estimation, I fix $\delta = 0.05$. To save computation costs, I consider a random subset of 200 of these possible deviations.[22]

[22] For $N = 3$, this procedure generates only 80 possible deviations. I consider all 80 deviations in this case.

The parameters $\{c, b\}$ enter the wholesaler profit function nonlinearly. Thus, a numerical method must be used to find the set of parameters described in equation 1.19. Following suggestions in (Bajari et al. 2007) and (Manski and Tamer 2002), I find the set Θ_s by minimizing $Q(\theta_s)$ via simulated annealing. I allow the algorithm to run for 25,000 iterations and then take the modal value as a point estimate.[23]

1.5.2.1 Identification

Given the estimation strategy above, it is important to understand what variation in the data identifies the parameters. The identification of the marginal cost and size distribution parameters are related to the shape of the price schedule itself. The wholesaler marginal cost c will be related to the lowest price option on the price schedule. As (Maskin and Riley 1984) point out, there is no efficiency loss under nonlinear pricing for consumers of the highest type. For a fully nonlinear schedule, that equates to the lowest price offered being equal to the marginal cost. With the discrete nature of the price schedules, there will still be some difference between the lowest price option and the marginal cost c. Thus, c must be recovered through the supply side model rather than directly inferred from the lowest marginal price offered.

The distribution parameter, b, is related to the curvature of the price schedule, i.e. the difference between the pricing options and where the cutoffs occur. To gain some intuition about this identification, consider the case of a degenerate distribution of type. In this case, retailers would be homogeneous and so no price discrimination is necessary and the optimal

[23]Strictly, this approach should return a set of vectors θ_s. In practice, most products have only one or two estimated vectors of coefficients that are near the minimum value of the objective value function. Running the algorithm for longer would likely generate additional points.

tariff would only have a single price. The closer the distribution is to being clustered around a single point, the smaller the steps between pricing options.

Figure 1.5 gives another intuitive look at how the supply side parameters impact the optimal price schedule. The baseline assumption (given by the blue line in the figure) is that the marginal cost is \$4 and F is uniform ($b = 1$). Increasing the marginal cost to \$8 while keeping the distribution the same shifts the entire price schedule up while keeping the cutoffs the same (as shown by the green line). Alternatively, keep the marginal cost the same and shifting to a more homogeneous distribution (as shown by the orange line with $b = 6$). While the prices also change for this shift, the lowest price offered is still approximately the same. Thus, price schedules with higher prices are associated with higher estimates of the marginal cost c. Likewise, price schedules with more closely spaced (and smaller magnitude) cutoffs are correlated with higher estimates of b (more homogeneous distributions).

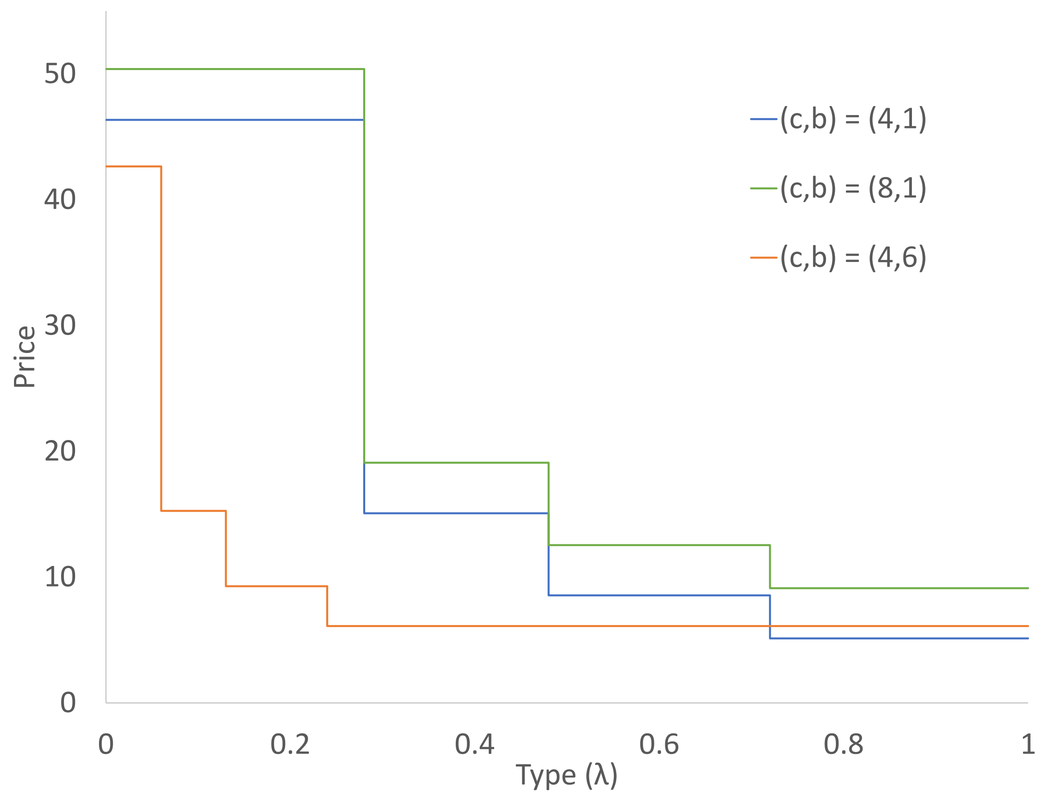
1.6 Results

Estimated coefficients for the consumer demand model are given first. Recovered parameters for the supply side model are presented second, as they are conditional on the estimated demand system.

1.6.1 Consumer Demand Estimates

Table 1.10 presents the estimated coefficients of the consumer utility function. The price coefficients are estimated in a first stage along with product dummies. These product dummies are then projected onto product characteristics to recover the characteristic parameters. Table 1.11 presents these second stage estimates. As expected, consumers prefer

Figure 1.5: Comparison of Optimal Price Schedules for Different Values of Supply Side Parameters.



higher proof imported liquor.

Table 1.10: Estimated First Stage Demand Coefficients

Variable	Coefficient	(SE)	p-value
Price	-.16	(0.00)	0.00
Price*ln($\frac{\text{Income}}{10000}$)	0.01	(0.00)	0.00
Product dummies	Y		
Log Likelihood	-205,263.63		

Critical to the model is consumer price response. This effect can be summarized by estimated demand elasticities for each product. Table 1.12 presents summary statistics for the estimated own price elasticities. Most of the estimated elasticities are above one in magnitude, suggesting demand for liquor is elastic. These estimates are consistent with the elasticity estimates presented in (Miravete et al. 2017) for Pennsylvania and (Conlon and Rao 2015) for Connecticut. As noted previously, the estimated specification does not control for correlation between prices and any unobserved product-month effects. However, given that the estimated elasticities are in line with those found in Connecticut and Pennsylvania, it appears that any endogeneity problem is of limited scope. Figure 1.6 presents distributions of the estimated elasticities by spirit type. Note that there is substantial variation in elasticities within a group. Further, it appears that the vodka category contains two types of products: high and low elasticity. This likely corresponds to basic and premium market segments.

1.6.2 Supply Side Estimates

I recover product specific supply side parameters for 1,161 product-months in 2011. Rather than focusing on the values for a particular product, I will present summary measures

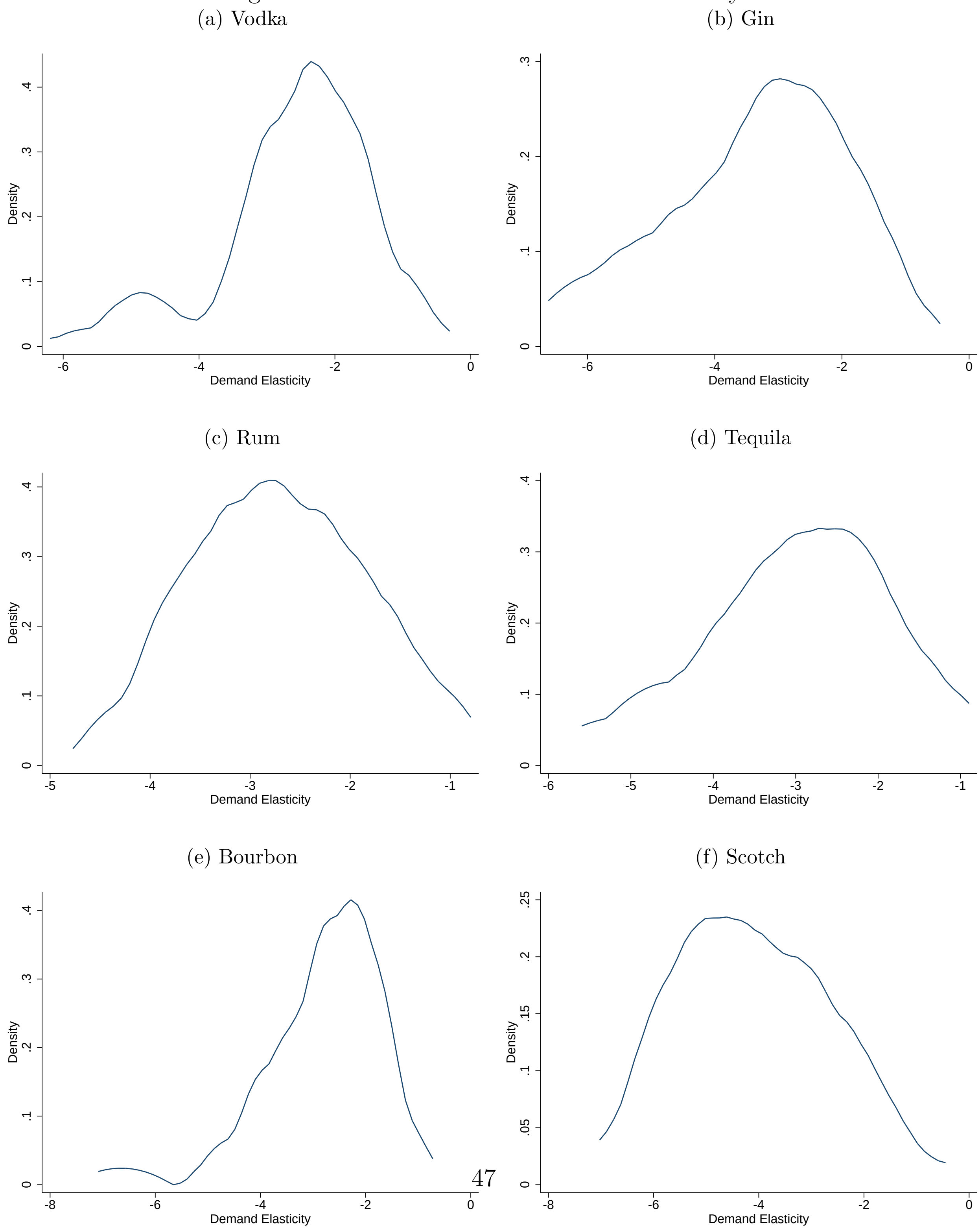
Table 1.11: Estimated Second Stage Demand Coefficients

Variable	Coefficient	(SE)	p-value
375ML	-1.14	(0.06)	0.00
750ML	Omitted	-	-
1L	0.37	(0.04)	0.00
1.75L	1.42	(0.04)	0.00
Imported	0.38	(0.03)	0.00
Proof	0.01	(0.00)	0.00
Gin	-0.90	(0.05)	0.00
Vodka	-0.32	(0.04)	0.00
Rum	-0.29	(0.05)	0.00
Scotch	0.17	(0.05)	0.0
Bourbon	0.42	(0.04)	0.00
Whiskey	-0.25	(0.05)	0.00
Tequila	-0.27	(0.09)	0.00
Other	Omitted	-	-
Constant	-3.95	(0.06)	0.00
Observations	22,479		
R^2	0.24		

Table 1.12: Estimated Own Price Elasticities

Mean:	-2.56
5th percentile:	-5.68
95th percentile:	-1.04
Number inelastic:	10/250

Figure 1.6: Distribution of Own Price Elasticity



of the parameters. This approach will help identify trends in the parameter values rather than focusing on outliers.

The supply side estimates reflect the nature of the observed price schedules. Many of the observed price schedules have relatively small jumps between the price options. This is reflected in relatively homogeneous distributions of retailer type. The cost parameters are below the lowest price offered on the price schedule, reflecting a positive markup, even for the largest purchase quantities.

Table 1.13 presents summary statistics for the recovered parameters.[24] The average recovered wholesaler marginal cost is \$14.61 per bottle. The distribution parameters are generally quite large, reflecting distributions with very little heterogeneity among retailers.[25] There is little correlation between the cost and distribution parameters.

Table 1.13: Distribution of Recovered Supply Side Parameters

Parameter	Mean[a]	SD[a]	Percentile[a]		
			25^{th}	50^{th}	75^{th}
Marginal Cost (c)	14.61	9.12	7.71	12.84	18.70
Distribution Parameter (b)	35.20	8.59	31.62	39.98	39.99

[a] The mean, standard deviation, and percentiles represent measurements across products. In principal, the moment inequality approach should return a set of estimates for each parameter. For the marginal cost c and distribution parameter b, only a point is returned in practice.

[24] These statistics exclude 101 of the 1,161 product-months for various reasons. Reasons for exclusion include not moving from the initial guess during estimation, extremely small cost estimates (generally recovered to be less than \$0.01 per bottle) and some extremely large estimates of welfare changes. These extremely large changes are two orders of magnitude larger than the next largest estimate. These outliers appear legitimate, but represent only two products for a total of 11 product-months.

[25] For computational reasons, an upper bound of 40 was imposed on the distribution parameters. Beyond that, the assumed distributions become approximately singular and are difficult to work with numerically.

1.7 Counterfactual Simulations

The primary purposes of developing and estimating the structural model presented above is to evaluate the welfare impact of banning quantity discounts at the wholesale level. To do so, the estimates presented in the last section are used to solve the model under the observed number of price options. Wholesaler profits, average retailer profits (weighted by type), and consumer surplus are calculated. The model is then solved again for only a single price ($N = 2$) and the same numbers are calculated. By comparing the two sets of welfare numbers, I can assess the impact of banning quantity discounts.

In general, removing quantity discounts at the wholesale level decreases overall welfare. The effects on wholesalers and retailers are in line with the effects of removing quantity discounts in a final goods market: the wholesaler loses profits; retailers benefit, though the impact depends on their type; and effects are smaller when the distribution is more homogeneous. The impact on consumers is negative and roughly the same magnitude as the change in wholesaler profits.[26] Table 1.14 presents a summary of these impacts across products in dollars per product per month.

Table 1.14: Distribution of Welfare Changes Across Products (Monthly dollars)

	Mean	SD	Percentile				
			5^{th}	25^{th}	50^{th}	75^{th}	95^{th}
Wholesaler Profit	-113	165	-408	-132	-56	-25	-3
Average Retailer Profit	75	112	2	17	37	88	267
Consumer Surplus	-113	164	-406	-132	-56	-25	-3
Total	-151	217	-544	-177	-75	-34	-4

[26]The same 101 products are excluded from the following tables.

Table 1.15 presents the same summary statistics in percentage change from the baseline of allowing quantity discounts:

Table 1.15: Distribution of Welfare Changes Across Products (Percent Change)

	Mean	SD	Percentile				
			5^{th}	25^{th}	50^{th}	75^{th}	95^{th}
Wholesaler Profit	-26.6	0.99	-28.2	-26.8	-26.5	-26.4	-23.7
Average Retailer Profit	31.4	1.71	28.1	30.7	32.1	32.5	32.9
Consumer Surplus	-26.5	0.99	-28.2	-26.7	-26.5	-26.4	-23.6
Total	-13.9	0.84	-15.7	-13.9	-13.8	-13.7	-11.8

The change in wholesaler profits declines in magnitude as the distribution becomes more homogeneous. This is exactly what one would expect in a final goods market - the more homogeneous the distribution of buyers, the less there is to be gained from engaging in second degree price discrimination. Figure 1.7 presents an estimated fit line for the relationship between the change in wholesaler profits and the distribution parameter (recall that higher values mean a more homogeneous distribution).

The change in wholesaler profits declines in magnitude as the marginal cost increases. Figure 1.8 shows the relationship. The magnitude of the loss declines up to a point than plateaus. This reflects the lower profit of higher marginal cost items both with and without quantity discounts.

The corresponding changes for retailers are generally smaller in magnitude and opposite in sign. This suggests that quantity discounts increase total profits in the intermediate tier of the industry. The wholesaler extracts more profit than is lost by the retailers by implementing the nonlinear pricing. There are, however, a few products for which the gain

Figure 1.7: Change in Wholesaler Profits versus b

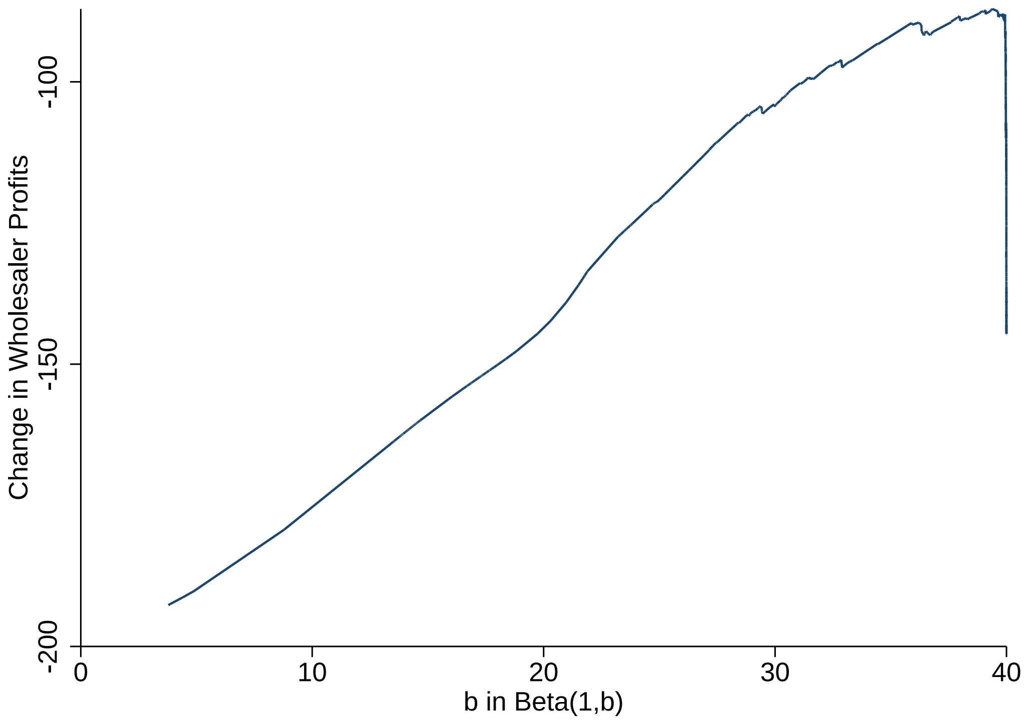
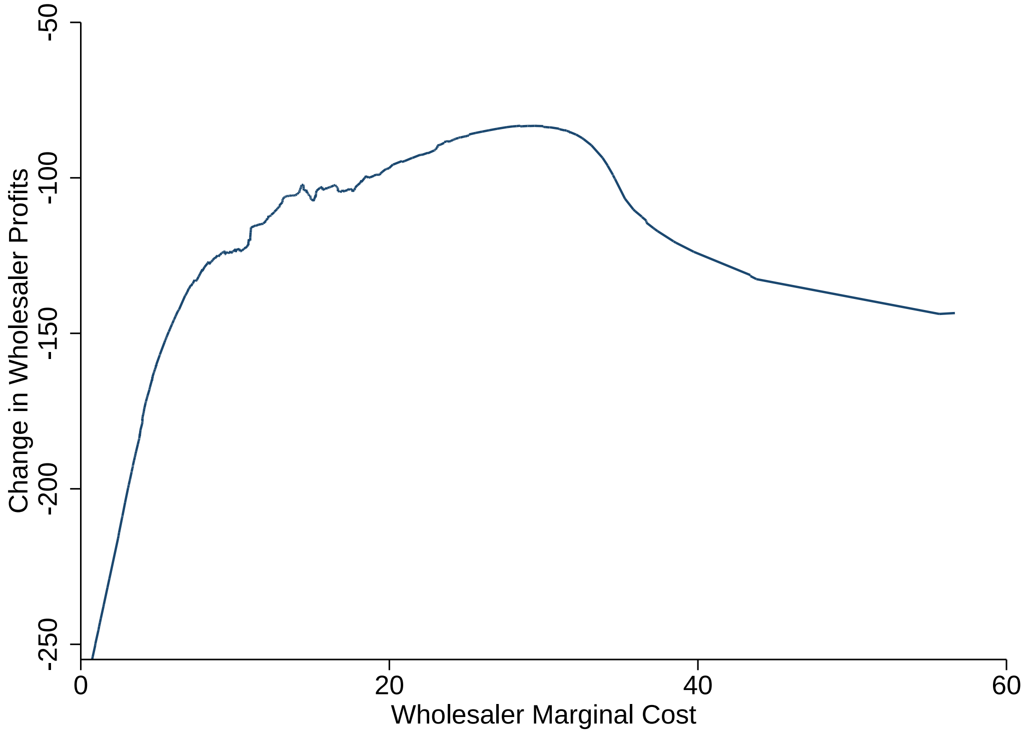


Figure 1.8: Change in Wholesaler Profits versus Marginal Cost

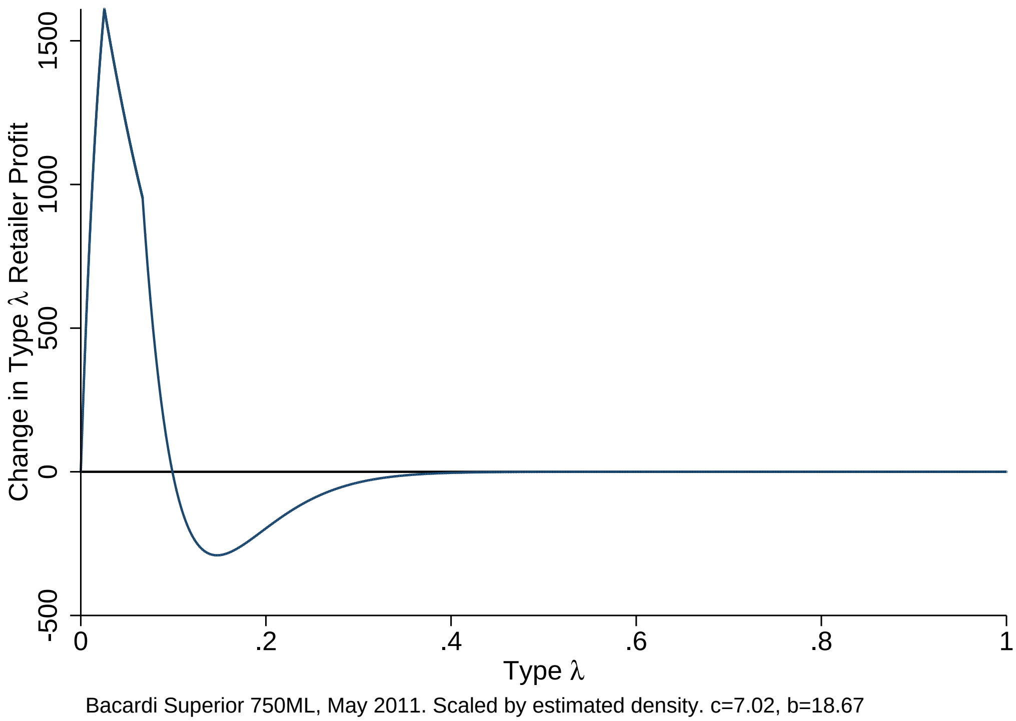


to retailers from removing quantity discounts is larger in magnitude suggesting that total profits increase by removing the price discrimination. These products have relatively heterogeneous estimated type distributions. However, there are only 55 of those products, representing only 5% of the sample once other outliers are excluded. Thus, it is difficult to draw strong conclusions from this subgroup.

While retailers see a profit increase on average, this hides heterogeneity among retailers. Retailers who were are the upper end of the type distribution and benefiting from quantity discounts see their input costs go up and thus profits decline. Figure 1.9 shows the change in profit as a function of retailer type for Bacardi Superior Rum, 750ML in May of 2011. The change in profits has been scaled by the estimated type density. Retailers at the very low end sell such small amounts that the profit changes must be small. However, as type increases the profit change first increases and then decreases. Large retailers see profit losses from the removal of quantity discounts. The average impact depends on the relative sizes of these groups. This can be seen by comparing the area under the curve with positive area representing gains to smaller retailers and negative area representing losses to larger retailers.

The impact on consumers is through changes in retail prices. Figure 1.10 shows how prices change with the removal of quantity discounts in the intermediate tier. The sales weighted average price increases. The retailers with higher local demand (i.e. high type) see an increase in costs and thus raise their prices. Smaller retailers do experience a decrease in input costs and so decrease their prices. However, the relative market share of these smaller retailers is not enough to outweigh the impact of the change of the large retailers. The net

Figure 1.9: Change in Retailer Profits by Type



effect is an increase in the price paid by consumers.[27]

Figure 1.10: Comparison of Average Retail Prices

Interestingly, the changes in consumer surplus show similar patterns to the changes in wholesaler profit. Figure 1.11 shows the relationship to the distributional parameter. Consumer surplus losses are smaller with more homogeneous distributions. Finally, Figure 1.12 shows that consumer surplus losses decline as wholesaler marginal costs increase.

The similar magnitudes and patterns of consumer surplus and wholesaler profits are

[27]The unweighted retail price does decrease. This would be the average price if one walked into several retailers and took a simple average of the prices. While this is reflective of the prices available to consumers, it is not reflective of that paid by consumers. The measurement difference is important if one were trying to assess the impact of a policy of banning quantity discounts directly.

Figure 1.11: Change in Consumer Surplus versus b

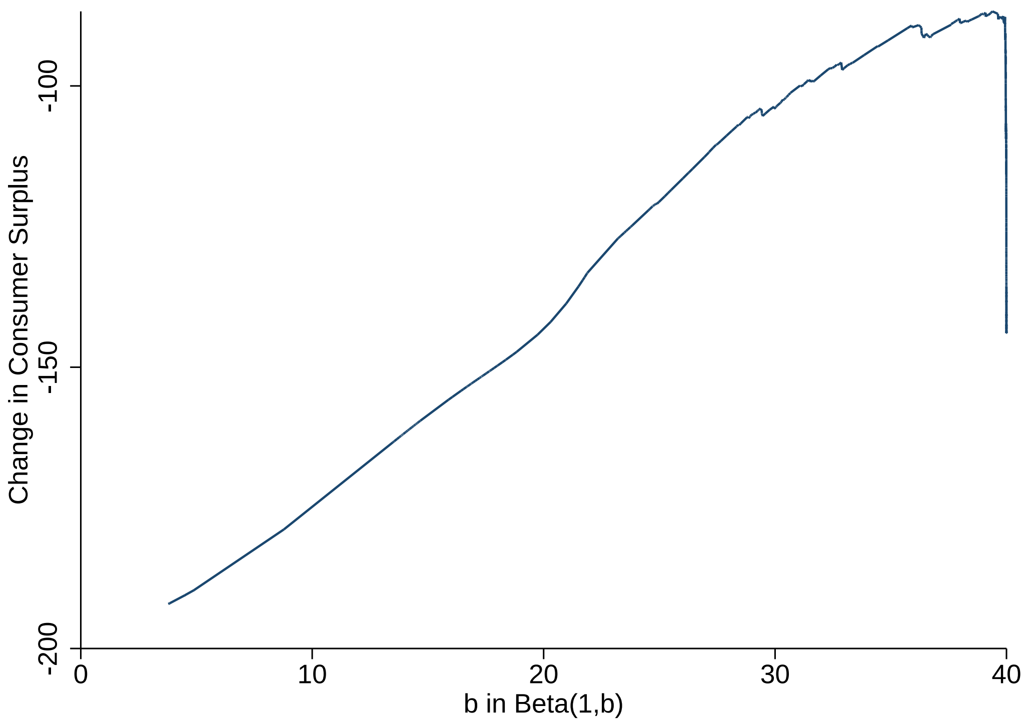
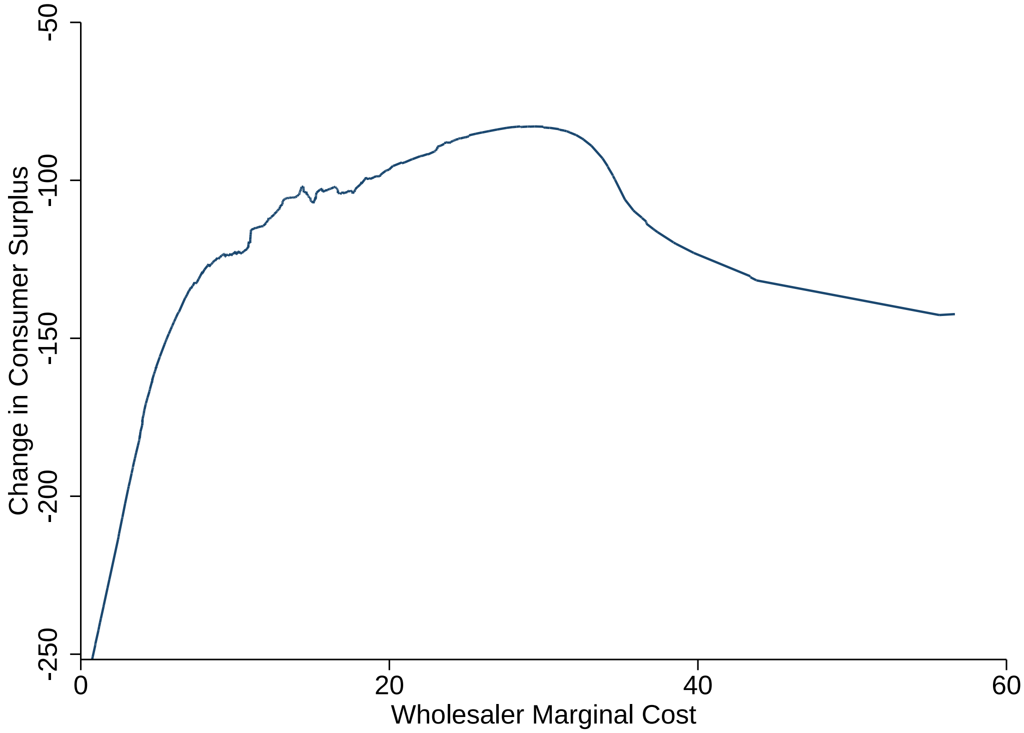


Figure 1.12: Change in Consumer Surplus versus Wholesaler Marginal Cost



interesting in that they are driven by the effects on different parts of the price schedule. That is, the losses in wholesaler profit are due to the reduced per-unit profit for smaller purchases on the lower quantity part of the price schedule and reduced purchases by large retailers. On the other hand, consumer surplus losses are driven by the increase in the price for the largest purchases at the high quantity part of the price schedule.

Given the similar patterns and magnitudes seen in wholesaler profit and consumer surplus, it is no surprise that total welfare exhibits similar patterns. Total welfare losses are smaller for more homogeneous distributions, higher cost products, and products with fewer parts in their price schedules. The graphs are omitted to avoid repetition.

The important conclusion to draw from this counter factual exercise is the different reasons for the welfare impacts in the two tiers. The story in the intermediate tier is similar to that seen in previous work on nonlinear pricing in a final goods market - wholesalers benefit from being able to price discriminate and charge high prices to smaller retailers. The welfare story in the retail tier is about which retailers consumers shop at. While some retailers see profit increases and lower their prices, the more popular retailers see input cost increases and profit losses due to increased prices. The retailers which benefited from quantity discounts are the more popular retailers - passing the discounts they receive on to consumers. Removing those discounts drives consumer prices up and thus consumer surplus down.

1.8 Conclusion

Theoretical predictions of the consequences of second degree price discrimination in intermediate markets are ambiguous. This paper empirically evaluated a particular case of quantity discounts in the wholesale liquor market in New York State. I developed and estimated a structural model of the optimal multi-part tariff which provides quantity discounts. The parameters of this model are then recovered from a unique data set of wholesale liquor price schedules. The recovered parameters are then used to evaluate a counter factual scenario and compare welfare with and without quantity discounts.

I find that quantity discounts typically increase total surplus in an intermediate goods context. Wholesalers benefit from being able to price discriminate against smaller purchasers. Larger retailers benefit from lower input prices. Consumers benefit because these larger retailers pass along the quantity discounts in the form of lower retail prices. Overall impacts are smaller with more homogeneous retailer markets, few numbers of price schedule parts, and higher wholesaler marginal cost.

Further work on understanding the relationship between retailer heterogeneity and welfare changes would be valuable. The products covered in this study are estimated to have relatively homogeneous distributions of retailers. It is difficult to extrapolate the conclusions in this paper to contexts with a more heterogeneous retail sector.

Finally, an extension to the current paper would be to include a more nuanced model of retailer competition over a longer horizon. The current paper considers the short term pricing impacts of removing quantity discounts with a stylized retailer model. Over a longer time horizon, one might be concerned about larger retailers leveraging their cost advantage to drive competitors out of the market. Another area where this paper could be extended is in the treatment of inventory. By ignoring inventory, the analysis precludes retailers from taking advantage of higher quantity discounts by grouping purchases over time. Inventory effects may be important for nonperishable products such as liquor.

Chapter 2

Foregone Profit from Using Multipart Price Schedules

2.1 Introduction

The use of nonlinear pricing by firms is an attempt to increase their profits over that afforded by simple single-price strategies when buyers are heterogeneous. For example, quantity discounts serve as a way for firms to price discriminate among their customers and extract additional rent. Economic theory predicts that the firm can extract maximum rents by offering a completely nonlinear price schedule where a unique (marginal) price is charged for every unit of the good sold. However, such complicated pricing strategies are rarely employed in practice. Thus, firms are foregoing profit by employing price schedules with a limited number of segments relative to the theoretical benchmark of a completely nonlinear price schedule. This paper seeks to evaluate the levels of foregone profit in a particular industry and relate those to product characteristics.

I examine the use of multi-part price schedules by liquor wholesalers in New York State. Heterogeneity in consumer demographics and preferences across retailers generates differences in demand across retailers. Given this downstream heterogeneity, wholesalers can increase their profits by engaging in price discrimination and offering multi-part price schedules.

These multi-part price schedules offer a series of price and quantity bands where price per unit is piece-wise constant within that quantity band. This feature is common to all the price schedules described in my data. These multi-part tariffs are common across industries, and examples can be readily found in industrial supplies,[1] office supplies,[2] and the financial sector.[3]

The model developed in Chapter 1 can be used to assess the foregone profit of not offering more segments. Wholesaler profit with the observed number of segments can be compared to the counterfactual profit with one additional segment to measure the foregone profit of one more segment. Further, the model can also provide an approximate value for the theoretical benchmark profit of a completely nonlinear price schedule.

In general, I find that foregone profits are small both in absolute terms and relative to the observed profit levels. In a similar vein, it appears that firms are capturing the majority of theoretically available profits. That is, there is little to be gained by offering a completely nonlinear price schedule. This suggests that firms are capturing much of the profits available despite using simpler strategies.

I find that these foregone profits are related to some product characteristics, such as spirit type. However, product characteristics have small effects in absolute terms. When considering percentages, product characteristics have no impact on foregone profit. This suggests that foregone profit depends more on the number of segments offered rather than particular product characteristics.

(Wilson 1993) §6.4 notes that, given the same set of buyers, multi-part price schedules

[1] www.fastenersuperstore.com/ offers multi-carton discounts for a wide array of fasteners and screws.
[2] www.uline.com offers per unit discounts on many common office, janitorial, and safety supplies.
[3] (Copeland and Garratt 2015) detail a nonlinear price schedule for settling financial accounts.

offer lower profits than the corresponding completely nonlinear price schedule. He indicates that the loss relative to the completely nonlinear schedule is on the order of the number of segments plus one squared and he goes on to show in §8.3 that this is a fairly general property of multi-part price schedules. However, the exact deviations depend on buyer preferences and the distribution of buyer heterogeneity. This paper seeks to complement this theoretical work by outlining which factors contribute to large or small deviations from the maximum obtainable profit.

Empirical work exploring approximations to optimal pricing strategies is relatively sparse. (Chu, Leslie and Sorensen 2011) examine approximations to optimal bundling strategies in a multi-product setting of pricing theater tickets. They find that the foregone profit from the simpler strategy is relatively small in percentage terms. (DellaVigna and Gentzkow 2017) note that retail chains employ uniform pricing over wide geographic areas. They argue that managerial costs prevent firms from employing the more complex (and likely more profitable) strategy of third degree price discrimination where different prices are charged at different stores. The current paper differs from this previous work in that it focuses on nonlinear pricing of a single good rather than bundling or geographical variation and examines many more products.

This paper is also related to work exploring the shape of nonlinear price schedules. (Busse and Rysman 2005) consider smooth approximations to observed multi-part price schedules and evaluate the effects of varying measures of competition on the shape of price schedules. This paper is similar in spirit, but examines a different measure of the shape of a price schedule.

2.2 Data

The data used to evaluate the foregone profit of offering a multi-part price schedule is the same as that discussed in Chapter 1. Foregone profit is a function of the number of segments offered. Therefore, the previous data discussion is supplemented with a discussion of the number of segments offered across price schedules.

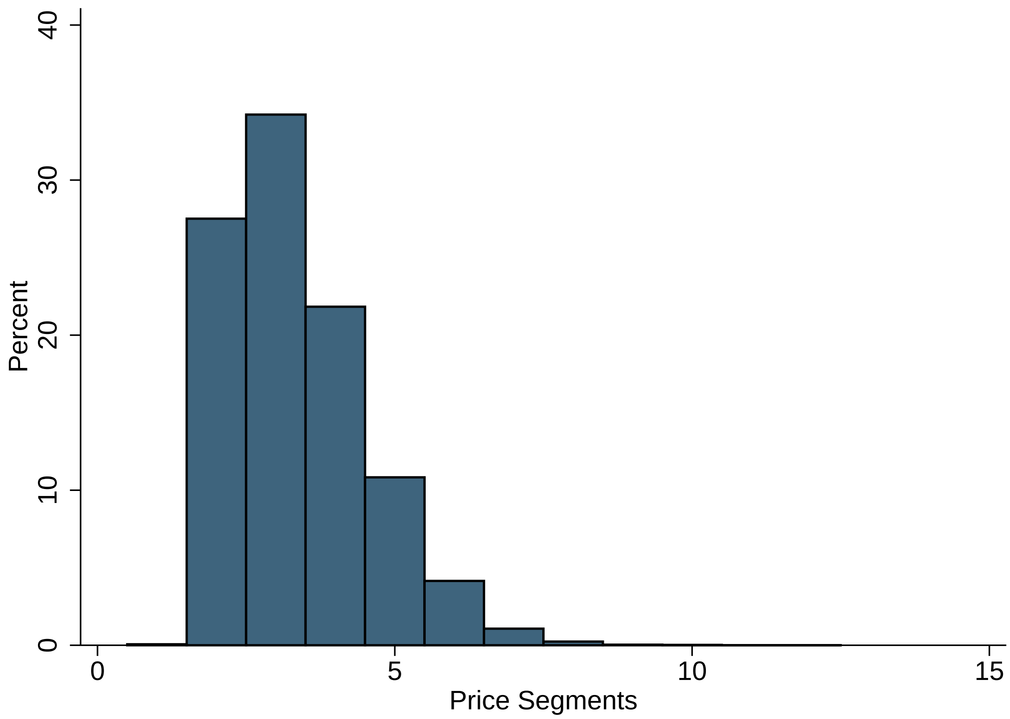
Figure 2.1 presents the unconditional distribution of the number of segments in a price schedule. There is substantial variation over the number of segments, with most schedules exhibiting between three and five segments. Almost no price schedules have one segment (i.e. the wholesaler does not offer a quantity discount). The relatively low number of segments offered is in line with theoretical predications from (Wilson 1993), suggesting that the observed schedules capture a large fraction of the profits potentially available.

The number of segments in the price schedules is unrelated to observable product characteristics. This suggests that wholesalers are not following rules such as "all gins have four prices offered", or "all 1.75L bottles have six prices." Rather, the wholesaler decides how many segments to offer on a product by product basis. Table 2.1 presents the results of a Poisson regression of the number of price schedule segments on some observable product characteristics derived from the price schedule data. These estimates are purely correlational, but serve as a useful way to summarize the variation. The coefficients are almost all precise zeros. The highest price, first quantity cutoff, and proof have little relation to the number of segments. The predicted number of segments for each spirit type is approximately the same as that for the baseline category (brandy). Likewise, bottle size has no effect, with 375ML, 750ML, and 1L bottles having approximately the same number of segments as 1.75L products.

Table 2.1: Poisson Regression of Number of Options on Product Characteristics

Variable	Coef.	SE	P-Value
First Price	-0.00	0.00	0.00
First Cutoff	-0.00	0.00	0.00
Proof	0.00	0.00	0.00
Gin	-0.11	0.01	0.00
Liqueur	-0.05	0.01	0.00
Rum	-0.04	0.01	0.00
Scotch/Whiskey/Bourbon	-0.06	0.01	0.00
Tequila	-0.16	0.01	0.00
Vodka	-0.09	0.01	0.00
1L	-0.10	0.00	0.00
375ML	-0.17	0.01	0.00
750ML	-0.15	0.00	0.00
Feb	0.05	0.01	0.00
Mar	0.04	0.01	0.00
Apr	0.03	0.01	0.00
May	0.04	0.01	0.00
Jun	0.06	0.01	0.00
Jul	-0.00	0.01	0.79
Aug	0.03	0.01	0.00
Sep	0.04	0.01	0.00
Oct	0.05	0.01	0.00
Nov	0.05	0.01	0.00
Dec	0.06	0.01	0.00
Constant	1.24	0.01	0.00
Observations	133,0034		
Pseudo R^2	0.01		

Figure 2.1: Distribution of the number of price options



The data description offered above suggests two conclusions about the foregone profit of multi-part price schedules in this industry. First, the wide variety in the number of segments offered suggests that the foregone profit may vary widely across products. Second, the fact that the number of segments is unrelated to observable product characteristics suggests that foregone profit might also be unrelated to observable product characteristics. That is, the foregone profit for a particular product may be product specific or related to market conditions.

2.3 Model

The structural model used is identical to that presented in Chapter 1. Therefore, rather than repeating the discussion presented previously, this section will provide intuition on what model components impact the foregone profit of offering only a finite number of segments. From the wholesaler's perspective, there are two key features of the model: the per unit profit from selling liquor, and the heterogeneity of the downstream buyers.

The wholesaler's per unit profit governs the overall profits from offering a particular price schedule. Retailers know consumer demand and make purchase decisions given a price schedule. The price schedule determines the revenue from the quantity sold. The wholesaler's cost structure determines the profit gained from that revenue. Combined, the derived demand and cost structure determine profits available at any particular number of price schedule segments. Recall that the model's cost structure is constant marginal cost while the derived demand structure is based on a logit demand system.

The heterogeneity of the downstream buyers (in this case retailers) determines how the wholesaler's profits change with the number of segments available. At one extreme, if buyers are homogeneous then the wholesaler cannot increase profits by offering more than a single price (i.e. a two-part tariff). If there are a finite number of buyer types then the wholesaler cannot increase profits beyond offering a different price to each type. The model assumes a continuum of buyer types and so the wholesaler can increase profits until a continuum of prices is offered (a fully nonlinear schedule). Put another way, buyer heterogeneity determines the curvature of profits with respect to the number of segments offered.

The goal of this paper is to understand the shape of wholesaler's profit with respect to the number of segments offered and how this shape varies across products. In the model, buyer heterogeneity is price schedule specific. The per unit profit function is common across products, aside from the wholesaler marginal cost which is price schedule specific. Thus, if there are patterns across products, this suggests that derived demand is relevant to the shape of wholesaler profits. Alternatively, if buyer heterogeneity is the primary factor driving the shape of foregone profits than we would not expect product characteristics to be particularly relevant.

2.4 Estimating Foregone Profit

Using the model to estimate foregone profit is straight forward. Model implied profits are calculated at three different number of segments: the observed number of segments, one additional segment, and at an infinite number of segments (corresponding to a completely nonlinear price schedule).[4] Profits can then be compared between the three different number of segments to gain an understanding of the foregone profits.

Formally, define the foregone profits between the observed number of segments and one additional segment as:

$$\underline{\zeta}_{N+1} = \pi_w(N+1) - \pi_w(N) \; , \tag{2.1}$$

where $\pi_w(N)$ is the wholesaler profit from offering $N-1$ segments (i.e. an N part tariff). While this value represents foregone profits, it also represents a lower bound on the cost of

[4]In practice, the profits at an infinite number of segments will be approximated with a finite but large number of segments. For all products, 15 segments was used as the approximation point. Thus, the values presented below strictly represent a lower bound on the foregone profits.

offering $N+1$ segments. That is, $\underline{\zeta}_{N+1}$ represents the minimum complexity cost necessary to rationalize the observed number of segments. This complexity cost might arise from managerial decision making costs, marketing costs, operational research costs to develop the price schedule, or other reasons.

A normalized value can also be considered to help account for different profit scaling across products. The normalized value is given by:

$$\underline{\zeta}^{pct}_{N+1} = \frac{\pi_w(N+1) - \pi_w(N)}{\pi_w(N)} \; . \tag{2.2}$$

Similarly, define the percent of maximum profits obtained as:

$$\zeta^{pct}_{\infty} = \frac{\pi_w(N)}{\pi_w(\infty)} \; , \tag{2.3}$$

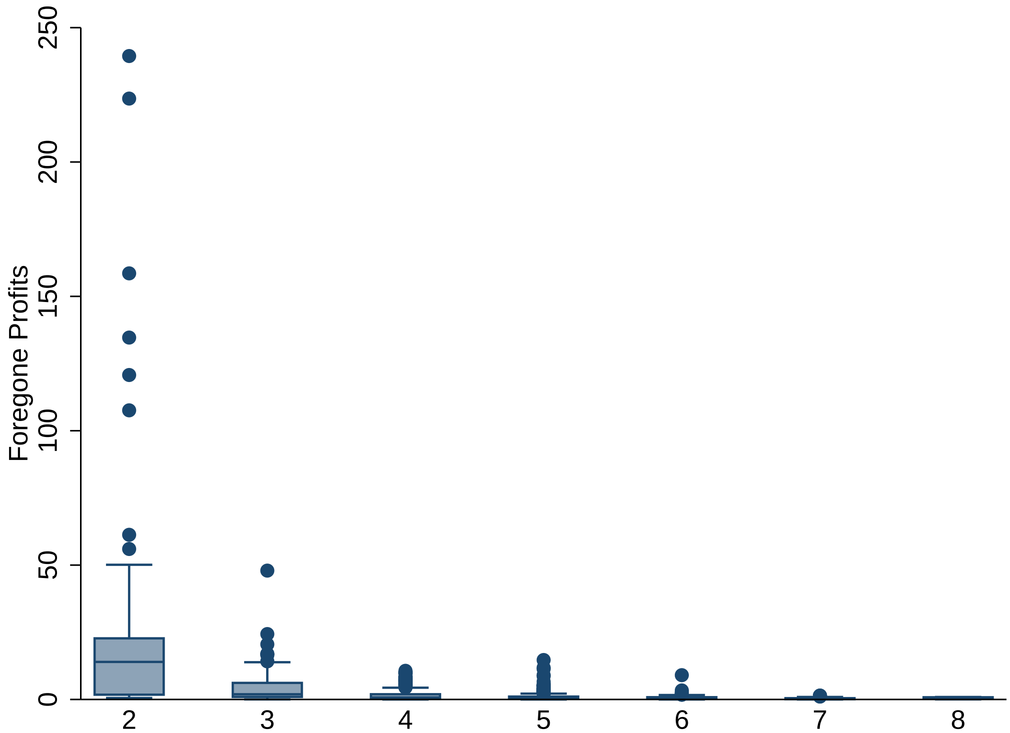
where $\pi_w(\infty)$ is the profit available from offering the completely nonlinear price schedule.

2.5 Results

The foregone profit was estimated for 1,060 product months over 2011.[5] Estimates of foregone profits are denoted in dollars per product per month and appear relatively small. Figure 2.2 presents summary statistics of the profit foregone profit by not offering an additional segment ($\underline{\zeta_{N+1}}$), by the total number of segments offered.

[5] As in Chapter 1, these statistics exclude 101 of the 1,161 product-months for various reasons. Reasons for exclusion include not moving from the initial guess during estimation, extremely small cost estimates (generally recovered to be less than $0.01 per bottle) and some extremely large estimates of welfare changes. These extremely large changes are two orders of magnitude larger than the next largest estimate. These outliers appear legitimate, but represent only two products for a total of 11 product-months.

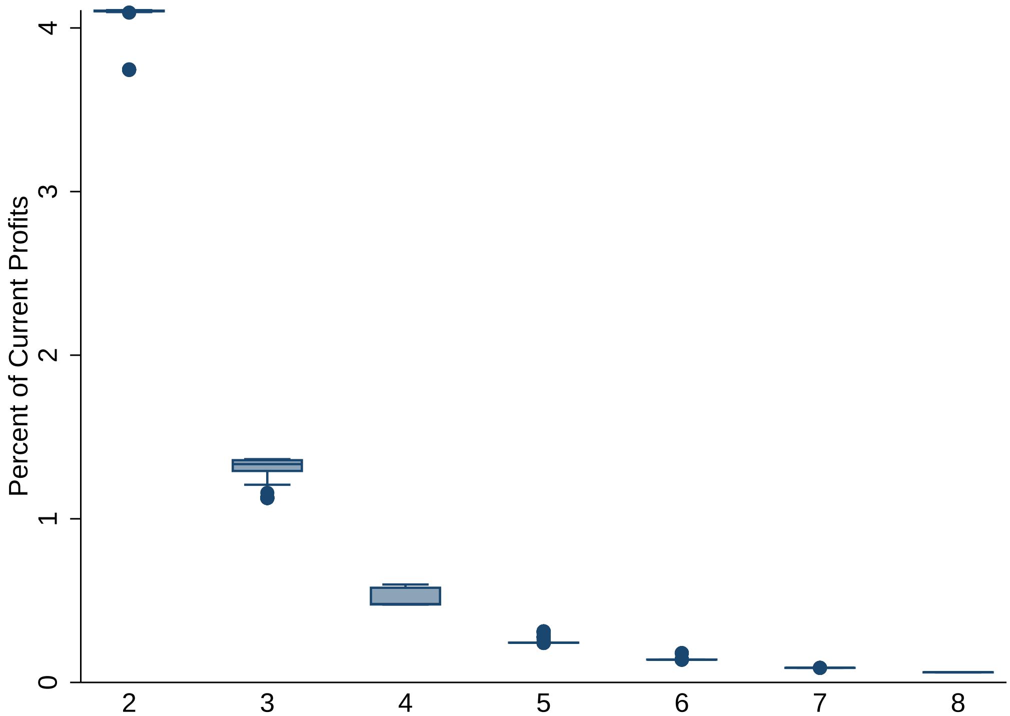
Figure 2.2: Distribution of Foregone Profit of Not Offering an Additional Segment



As expected, the foregone profit of adding an additional segment generally decreases as the segments offered increases. Note the dramatic decrease in level of foregone profits between offering two and three prices. This suggests that the majority of the profits available form a nonlinear price schedule can be obtained from just two prices. Further, there is substantial variability within a total number of segments. This suggests that product specific factors are likely important.

Figure 2.3 presents the normalized value of the foregone profits ($\zeta^{pct}_{\underline{N}+1}$). Recall that this is the percentage increase in current profits are available by offering an additional segment.

Figure 2.3: Distribution of Foregone Profit of Not Offering an Additional Segment



In percentage terms, the foregone profit is now much less widely dispersed. It is also extremely small. The median product offering two prices could increase profits by less than 5% by offering an additional price. This rapidly drops to approximately 1.5% when offering three prices. Considering foregone profit in percentage terms corrects for different profit scaling between products. Thus, the focus here is primarily on the curvature of profits with respect to the number of segments rather than the overall level.

Relating these graphs back to the model indicates that variation in the per unit profit across products is more important than variation in the buyer heterogeneity across products. The foregone profit looks rather widely dispersed across products. However, once product specific scaling is controlled for by considering foregone profits as a percent, products look very similar to each other.

It appears that product characteristics play some role in the foregone profits. Table 2.2 presents the results of regressing both the level and percent of foregone profits on product characteristics and time dummies.

Which product characteristics appear important depends on whether levels or percentages are considered. For example, being a gin product appears to matter for levels but not percentages. Interestingly, the percent regression explains almost all of the variation in the data, suggesting that the curvature of profits with respect to the number of segments is unrelated to product characteristics and is similar across products.

Comparing the profits at the observed price schedule to the maximum obtainable profits reveals similar patterns. Recall that the maximum profits are achieved by using a completely nonlinear schedule. The observed schedules all offer a finite number of parts and so achieve only a portion of the total profits available.

Figure 2.4 plots the distribution of the percentage of maximum available profits over the number of segments offered. Offering two prices obtains almost all of the profit available. Almost 98% of the profits can be obtained with just three prices. These values are much higher than those predicted by the theory.[6] That is, simple pricing strategies obtain more of the available profit in practice than suggested by theory.

[6]Recall that (Wilson 1993) suggests that foregone profits are on the order of the number of segments squared. That approximation suggests that three prices would capture 88% of profits.

Table 2.2: Regression of Foregone Profit on Product Characteristics

Dependent Variable:	Level	Percent
Proof	-0.011	-0.000
	(0.020)	(0.000)
Gin	-2.141	0.006
	(0.958)	(0.005)
Vodka	-0.906	-0.004
	(0.679)	(0.004)
Rum	-4.429	-0.031
	(1.537)	(0.009)
Scotch	-1.442	0.012
	(0.847)	(0.005)
Bourbon	3.920	-0.007
	(1.936)	(0.004)
Whiskey	-1.743	-0.005
	(0.696)	(0.005)
Tequila	1.081	0.007
	(1.293)	(0.009)
750ML	2.859	-0.024
	(2.337)	(0.048)
1L	2.460	-0.017
	(2.419)	(0.135)
1.75L	5.220	-0.027
	(2.204)	(0.011)
Constant	28.850	4.119
	(8.599)	(0.020)
Month Controls	X	X
Total Segment Controls	X	X
Observations	1060	1060
R^2	0.259	0.998

Figure 2.4: Distribution of Percent of Maximum Profit Obtained

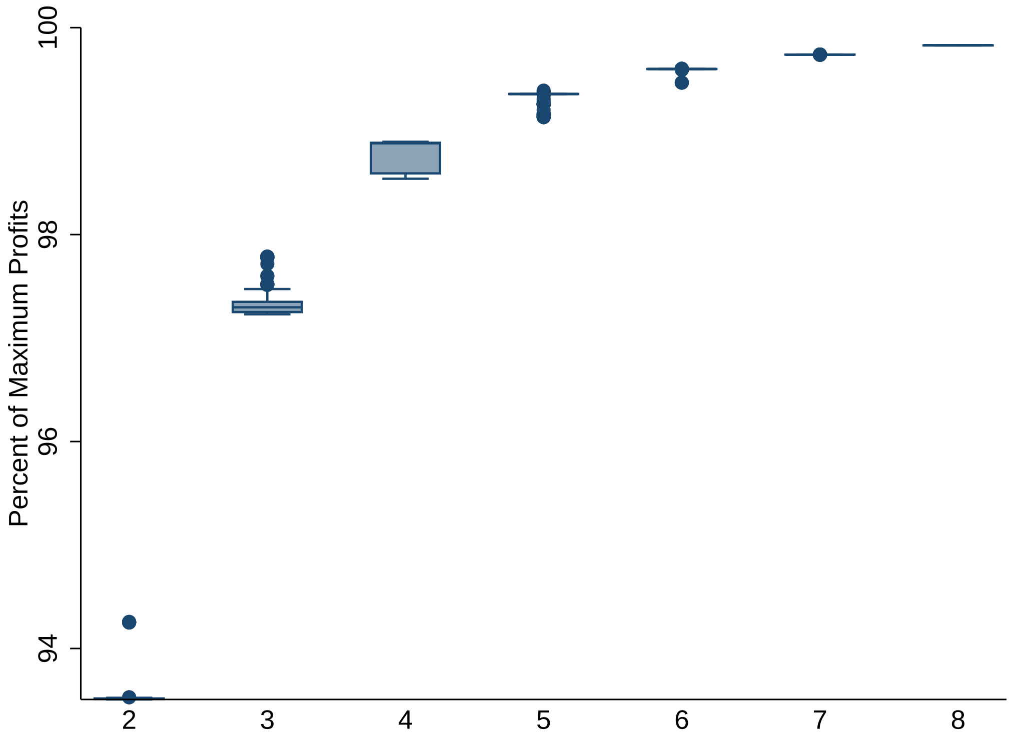


Table 2.3 relates the percent of maximum profits to product characteristics. Product characteristics appear to have little meaningful effect on the percent of profits obtained. Further, the simple model captures almost all the variation in the data suggesting that the percentage of profits obtained depends on the number of segments offered rather than particular product characteristics.

2.6 Conclusion

In practice, firms appear not to use relatively complicated pricing strategies. Rather, they rely on pricing strategies that are simpler, but offer lower profits than the theoretical

Table 2.3: Regression of Percent Maximum Profit on Product Characteristics

Dependent Variable:	Percent
Proof	0.001
	(0.000)
Gin	-0.017
	(0.014)
Vodka	0.010
	(0.009)
Rum	0.064
	(0.019)
Scotch	-0.038
	(0.014)
Bourbon	-0.022
	(0.010)
Whiskey	0.011
	(0.014)
Tequila	-0.015
	(0.022)
750ML	0.037
	(0.028)
1L	0.023
	(0.027)
1.75L	0.044
	(0.026)
Constant	93.477
	(0.046)
Month Controls	X
Total Segment Controls	X
Observations	1060
R^2	0.994

optimum. This paper evaluates these forgone profits in the context of multi-part price schedules in the New York liquor market.

This paper finds that foregone profits are small both in absolute terms and relative to the observed profit levels. This suggests that firms are capturing much of profits available despite using relatively simple strategies. It also appears that some product characteristics are related to the foregone profits, such as spirit type. However, product characteristics have small effects on the foregone profits.

Likewise, it appears that firms are capturing the majority of the profit available. That is, there is little to be gained by offering a complex nonlinear price schedule. The percentages obtained by a particular number of segments appear higher than those suggested by theory. This suggests that the gains from complex strategies might be lower than previously thought.

Evaluating complex pricing strategies in other product markets would be valuable future work. Further, trying to understand why the gains from complex pricing appear relatively invariant to the product considered would be interesting. While nominal values appear to vary across products somewhat, in percentage terms there is relatively little variation. This suggests that the approximate optimality of these simpler pricing strategies is a more fundamental insight.

Chapter 3

All Unit Discounts as a Price Discrimination Tool

3.1 Introduction

Nonlinear pricing takes many forms. Common types of nonlinear pricing include quantity discounts, bundling, and free allocations. Each of these particular forms of nonlinear pricing may have different sub-types that differ in their implementation. Different implementations of a nonlinear pricing scheme may then have different outcomes for market participants. That is, particular forms of nonlinear pricing might be better suited to particular goals. This paper is concerned with comparing two different forms of quantity discounts and considering how well each functions as a price discrimination mechanism.

Quantity discounts are generally characterized by marginal prices that decline as purchase size grows. There are two main forms of quantity discounts - incremental unit discounts and all-unit discounts. With incremental unit discounts, purchasers pay the marginal price listed for each quantity size up to their desired amount. That is, the quantity discount is applied only to the subsequently purchased units. All-unit discounts apply the discount to all previously purchased units. That is, when the threshold for a lower marginal price is reached, the purchaser saves on all previously purchased units as well.

On the surface, these two discount types are relatively similar. Both are characterized by declining marginal price schedules. However, the different implementations offer different incentives to purchasers. For incremental discounts, buyers only need consider whether the next marginal unit offers more marginal benefit than the marginal cost. With all-unit discounts, the buyer must consider not only whether that next unit has more benefits than costs but also whether the next marginal unit offers sufficient savings on previous units to warrant a purchase. The two incentive schemes will generate different observable patterns of purchases.

In this paper I characterize the optimal all-unit discount price schedule and compare it to the optimal incremental discount schedule. I then compare profits and price schedules.

All-unit discounts result in lower profits for sellers. Further, all-unit discounts result in higher marginal prices and lower purchase quantities in equilibrium. All-unit discounts better approximate incremental discounts when consumers have lower demand on average and are less heterogeneous. Finally, all-unit discounts generate bunching of consumer purchases in equilibrium posing challenges for estimation of the model.

Relatively little of the literature has considered all-unit discounts directly. As a form of quantity discounts, and thus a form of second degree price discrimination, this discussion of all-unit discounts is related to previous theoretical literature focusing on nonlinear pricing. Early theoretical work is largely concerned with finding the optimal nonlinear schedule among a broad class of pricing strategies. (Spence 1977), (Mussa and Rosen 1978), (Katz 1983), and (Maskin and Riley 1984) all take such an approach. (Wilson 1993) notes that, under certain regularity conditions, these completely nonlinear price schedules can be expressed as a series of two part tariffs. Such a sequence of two part tariffs is equivalent to the incremental discounts discussed above. However, none of these previous authors have considered implementation of second degree price discrimination through the use of all-unit discounts. (Amornpetchkul 2017) notes that all-unit discounts generate different purchase behavior than incremental discounts, but considers only a simplified two-price version of each discount scheme.

Another branch of literature considers all-unit discounts as a method of foreclosing competition. (Kolay Sreya, Shaffer Greg and Ordover Janusz A. 2004) present a discussion showing that while all-unit discounts have been considered a foreclosure mechanism, they can arise in equilibrium without that motive. They go on to show that, relative to incremental discounts, all-unit discounts might increase or decrease welfare in a double marginalization context. (Conlon and Mortimer 2013) examine the exclusionary nature of all-unit discounts in an empirical setting. They find that wholesaler's offering all-unit discounts can cause retailers to not carry competitors' products. The current paper differs in that it considers all unit discounts as a price discrimination mechanism rather than a foreclosure mechanism.

3.2 Intuition

The implementation of quantity discounts as all-unit discounts or incremental discounts generates different incentive schemes for buyers. This section tries to outline how those incentive schemes differ and what the observable consequences of those differences are. In both instances price schedules offering a finite number of segments will be considered. That is, price schedules are given by a series of marginal prices and quantity thresholds. Finally, there are a large number of consumers indexed by some continuous "type" θ such that higher types purchase more at any given point.[1]

[1]This is essentially assuming single crossing property detailed in (Maskin and Riley 1984) and discussed previously in Chapter 1.

The primary difference in the incentives between all-unit discounts and incremental discounts is the shape of the total payment schedules. For incremental discounts, the total payment as a function of units purchased is always increasing. Recall that incremental discounts only apply to marginal units, not previously purchased units. Thus, the equilibrium price schedule is continuous and increasing in quantity. For all-unit discounts total payment as a function of quantity purchase decreases at the quantity thresholds because the new marginal price applies to all previous units. Thus, the total payment function will have discontinuities at the quantity thresholds.

Figure 3.1 shows the total payment functions for a hypothetical set of incremental discounts and all unit discounts. The marginal prices in both cases are the same, as are the quantity cutoffs.[2] Note that the all unit discount schedule contains the discontinuities where the total payment decreases by purchasing one more unit at the quantity thresholds. This is the discontinuity discussed above.

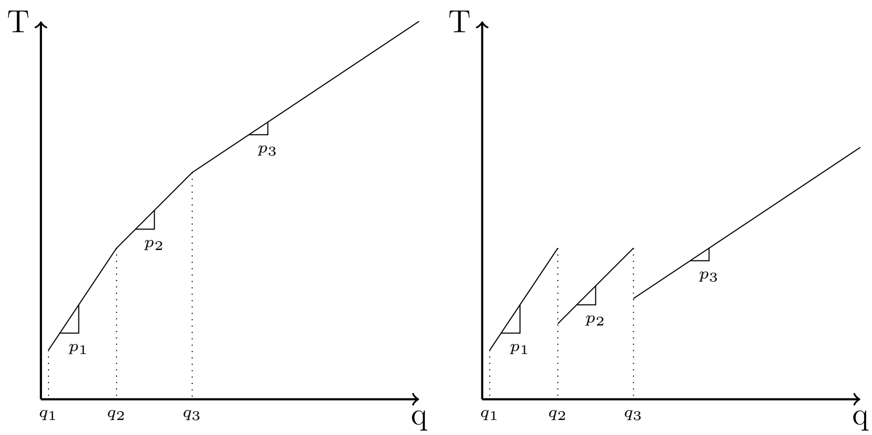
Also notice that the total payment is lower with all-unit discounts with the same marginal prices. Thus, one might expect that the marginal prices would be higher under all-unit discounts so that the seller captures some of the profit implied in the fixed fee through the marginal prices. Thus, equilibrium purchases should generally be lower under all-unit discounts than under incremental discounts.

Given the discontinuity in the total payment schedule, some buyers could actually lower their total payment by purchasing more units. For these consumers, the reduction in total cost and marginal benefit of the additional units increases their total payoff. That is,

[2] Cutoff q_1 corresponds to the minimum purchase size the seller is willing to facilitate in both cases. In the incremental discount case this is equivalent to setting a fixed fee in the first segment.

Figure 3.1: Total Payment Function

(a) Incremental Discounts (b) All Unit Discounts



for customers who optimally purchase some amount at p_1 that is "close enough" to q_2, they are willing to purchase additional units beyond their optimal amount to lower their overall cost.

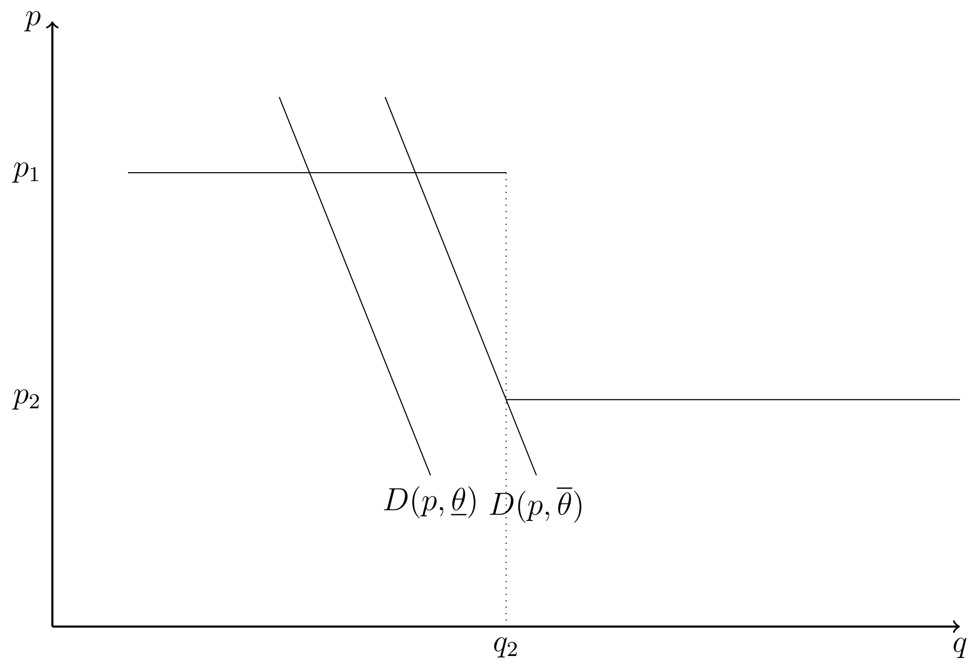
Figure 3.2 depicts just such a consumer.[3] Consumer $\underline{\theta}$ is indifferent between buying $D(p_1, \underline{\theta})$ at p_1 and q_2 at p_2, where $D(p, \theta)$ is the optimal purchase of type θ at price p. Note that $D(p_2, \underline{\theta}) < q_2$ implying that this consumer wants to purchase less than q_2 even at the lower marginal price. However, the cost savings is high enough by moving from p_1 to p_2 so that they prefer buying the "extra" units.

Figure 3.2 also shows the first consumer who's optimal purchase is q_2, i.e. type $\overline{\theta}$. This consumer can get more for a lower total payment at p_2. However, it is important to note that q_2 is the optimal purchase size for this consumer so there are no "extra" units.

Together types $\underline{\theta}$ and $\overline{\theta}$ define a range of types who all purchase exactly q_2. For almost

[3]I would like to thank David Sibley for helping me to clarify the problem with a similar figure.

Figure 3.2: Bunching of Consumes



all of these types, they are buying more than is implied by their demand function. This behavior will generate bunching of purchases at q_2. That is, in the distribution of purchase sizes there will be a point mass at q_2. Such behavior is induced by any discontinuities in the price schedule, thus there may be multiple bunch points in the purchase distribution.

This bunching distinguishes the purchase behavior under all-unit discounts from that seen in incremental discounts. There is no such cost savings for buying more under incremental discounts, and so there will be no bunching in the purchase distribution at the cutoff quantities. The seller will need to consider the bunching when designing the optimal all-unit discount schedule.

3.3 Model

This section describes the seller problems that define the optimal all-unit discount and incremental discount schedules. The solution to these problems describe the equilibrium price schedules given a consumer population.

In both cases, sellers face a heterogeneous distribution of buyers characterized by some type $\theta \sim F(\theta)$. Without loss of generality, I will assume F has support on $[0, 1]$. Consumers have some utility function:

$$V(\theta, q, p) = u(\theta, q) - P(q) \ , \tag{3.1}$$

such that $P(q)$ is the total payment for buying quantity q and consumers with higher θ have higher willingness to pay for a given quantity. Finally, assume the seller faces a constant marginal cost. The seller's problem is then to find the profit maximizing $P(q)$ given consumer preferences and their marginal cost.

3.3.1 Incremental Discounts

For a seller offering incremental discounts, their price schedule can be characterized by a series of fixed fees and associated marginal prices (A_k, p_k) so that the price schedule is given by the lower envelope of the two part tariffs. Quantity cutoffs are defined by where the applicable two part tariff changes. Formally, the seller picks a piecewise linear function $P(q)$ where each segment is given by:

$$P_k(q) = A_k + p_k q \ . \tag{3.2}$$

To enforce continuity of the price schedule, the seller is constrained in the choice of fixed fees. As is typical in a price discrimination problem, the seller will want to ensure that the lowest valuation buyer in a particular segment is indifferent between that segment and the previous.[4] Formally, this gives a set of conditions:

$$u(\theta_k, D(p_k, \theta_k)) - A_k - p_k D(p_k, \theta_k) = u(\theta_k, D(p_{k-1}, \theta_k)) - A_{k-1} - p_{k-1} D(p_{k-1}, \theta_k) \ , \quad (3.3)$$

where θ_k is the minimum type buying in segment k, i.e. the cutoff type.

These conditions can be used to express the seller's problem entirely in terms of prices and cutoff types. For a given number of prices $N-1$:

$$\begin{aligned} \max_{p_k, \theta_k|_{k=1}^{N-1}} \sum_{k=1}^{N-1} \Bigg[\int_{\theta_k}^{\theta_{k+1}} & (p_k - c) D(\theta, p_k) dF(\theta) \\ & + (1 - F(\lambda_k))(u(\theta_k, D(\theta_k, p_k)) - p_k D(\theta_k, p_k) \\ & \qquad -u(\theta_k, D(\theta_k, p_{k-1})) + p_{k-1} D(\theta_k, p_{k-1})) \Bigg] \ . \quad (3.4) \end{aligned}$$

The integral term represents the per unit profit from buyers in segment k. The second term defines the increment in the fixed fees between segments. Implicitly, all purchasers above a particular segment are paying the incremental fixed fee for that segment.

There are also some relevant boundary conditions. At the upper end, $\theta_N = 1$ representing the highest valuation consumer. At the lower end, p_0 represents not purchasing such that $D(\theta, p_0) = 0$.

The optimal price schedule is characterized by the first order conditions of this problem. There is one first order condition for each price and each cutoff type giving a total of

[4] This discussion is almost identical to that seen in Chapter 1.

$2(N-1)$ conditions. The optimality condition for a price p_k is given by:

$$\int_{\theta_k}^{\theta_{k+1}} [(p_k - c)D_p(\theta, p_k) + D(\theta, p_k)]\, dF(\theta)$$
$$+ (1 - F(\lambda_k))\,(u_q(\theta_k, D(\theta_k, p_k))D_p(\theta_k, p_k) - p_k D_p(\theta_k, p_k) - D(\theta_k, p_k))$$
$$- (1 - F(\lambda_{k+1}))\,(u_q(\theta_{k+1}, D(\theta_{k+1}, p_k))D_p(\theta_{k+1}, p_k) - p_k D_p(\theta_{k+1}, p_k) - D(\theta_{k+1}, p_k))\ . \quad (3.5)$$

The optimality condition for the cutoff type θ_k is[5]:

$$- (p_k - c)D(\theta_k, p_k)f(\theta_k) + (p_{k-1} - c)D(\theta_k, p_{k-1})f(\theta_k)$$
$$+ (1 - F(\lambda_k))\,[u_\theta(\theta_k, D(\theta_k, p_k)) - u_\theta(\theta_k, D(\theta_k, p_{k-1}))]$$
$$- f(\theta_k)\,[u(\theta_k, D(\theta_k, p_k)) - p_k D(\theta_k, p_k) - u(\theta_k, D(\theta_k, p_{k-1})) + p_{k-1}D(\theta_k, p_{k-1})]\ . \quad (3.6)$$

Together, these first order conditions form a system of (potentially nonlinear) equations that can be solved for the optimal price schedule. Substituting the equilibrium prices and cutoff types into the consumer indifference conditions defines the associated fixed fees.

3.3.2 All-Unit Discounts

For a seller offering all-unit discounts, the price schedule is defined by a series of prices and quantity cutoffs. Formally, the seller picks a pricing function $P(q)$ where each segment is given by:

$$P_k(q) = p_k q\ . \quad (3.7)$$

[5]The optimality condition has been simplified by imposing the equilibrium purchase condition of consumers such that $u_q(\theta, p) = p$.

Recall that all-unit discounts generate a discontinuous total payment function and one defining feature is that there is bunching at the cutoff quantities. Therefore, one must define the boundaries of the set of consumers who bunch.

The lower bound on the bunching region is the consumer who is indifferent between buying in the current segment or purchasing exactly at the quantity cutoff. Formally, for a particular price p_k and associated cutoff quantity q_k, the lower bound on the bunching region $\underline{\theta}_k$ is defined by:

$$u(\underline{\theta}_k, D(\underline{\theta}_k, p_{k-1})) - p_{k-1}D(\underline{\theta}_k, p_{k-1}) = u(\underline{\theta}_k, q_k) - p_k q_k \ . \tag{3.8}$$

The upper bound on the bunching region is defined as the consumer who's optimal purchase at p_k is exactly q_k. Formally, $\overline{\theta}_k$ is given by:

$$D(\overline{\theta}_k, p_k) = q_k \ . \tag{3.9}$$

All consumers with type $\theta \in \left[\underline{\theta}_k, \overline{\theta}_k\right]$ purchase exactly q_k. These conditions define a bunching region for each segment.

Given these boundaries, the seller's problem is given by:

$$\begin{aligned} \max_{p_k, q_k|_{k=1}^{N-1}} \sum_{k=1}^{N-1} \int_{\underline{\theta}_k}^{\overline{\theta}_k} (p_k - c) q_k dF(\theta) + \int_{\overline{\theta}_k}^{\underline{\theta}_{k+1}} (p_k - c) D(\theta, p_k) dF(\theta) \ , \qquad (3.10) \\ \text{s.t.} \quad u(\underline{\theta}_k, D(\underline{\theta}_k, p_{k-1})) - p_{k-1}D(\underline{\theta}_k, p_{k-1}) = u(\underline{\theta}_k, q_k) - p_k q_k \ , \\ D(\overline{\theta}_k, p_k) = q_k \ , \end{aligned}$$

and the boundary condition that $\underline{\theta}_N = 1$. The constraints can be substituted into the profit function to express the problem as an unconstrained problem. However, that approach relies on being able to express the boundary types in closed forms.

The first order conditions for this problem characterize the optimal all-unit discount price schedule. The optimality condition for prices p_k are given by:

$$
\begin{aligned}
- (p_k - c) q_k f(\underline{\theta}_k) \frac{\partial \underline{\theta}_k}{\partial p_k} + \int_{\underline{\theta}_k}^{\overline{\theta}_k} q_k dF(\theta) + (p_k - c) D(\underline{\theta}_{k+1}, p_k) f(\underline{\theta}_{k+1}) \frac{\underline{\theta}_{k+1}}{\partial p_k} \\
+ \int_{\underline{\theta}_{k+1}}^{\overline{\theta}_k} \left[(p_k - c) D_p(\theta, p_k) + D(\theta, p_k) \right] dF(\theta) = 0 \ . \quad (3.11)
\end{aligned}
$$

The optimality conditions for the cutoff quantities are given by:

$$
-(p_k - c) f(\underline{\theta}_k) \frac{\partial \underline{\theta}_k}{\partial q_k} + \int_{\underline{\theta}_k}^{\overline{\theta}_k} (p_k - c) dF(\theta) = 0 \ . \quad (3.12)
$$

Combined, this gives a system of $2(N-1)$ nonlinear equations that can be solved for the optimal price schedule.

3.4 Comparison

This section calculates the optimal incremental discount schedule and optimal all-unit discount schedule given the same set of consumers. Given the price schedules, I can compare prices, profits, and consumer surplus to understand how the implementation of quantity discounts matters.

Particular functional forms need to be selected to implement the above models. In particular, a marginal cost c, type distribution F, and consumer demand D must be chosen.[6].

[6]Choosing a demand function implies a utility function for consumers. Alternatively, one could pick a consumer utility function which implies a particular demand.

For the purposes of this exercise:

$$c = 0\,, \tag{3.13}$$

$$F(\theta; b) = 1 - (1-\theta)^b\,, \tag{3.14}$$

$$D(\theta, p) = \theta(1-p)\,. \tag{3.15}$$

The consumer type distribution is depends on another parameter b and is equivalent to assuming that types are distributed according to a Beta$(1, b)$ distribution. This parameter governs the heterogeneity of the type distribution. Higher values of b shift probability mass towards zero and correspond to lower average type as well as lower variance in the distribution.

Given this parameterization, the optimal all-unit and incremental discount schedules can be calculated. To better understand how the all-unit discount schedules correspond to the incremental discount schedules, different number of segments (N) and different type distributions will be considered. Specifically, $N \in [2, 6]$ and $b \in \{0.5, 1.0, 2.0\}$ are examined. This translates to between 1 and 5 prices and mean types of $\frac{1}{3}, \frac{1}{2}$, and $\frac{2}{3}$ respectively.

Table 3.1 presents the relative profit for implementing an all-unit discount schedule versus an incremental discount schedule. In general, all-unit discounts offer lower levels of profit than the incremental discount schedule. In a sense, the all-unit discount schedule is a restricted form of the incremental discount schedule where the fixed fees have been constrained to be zero for all segments. Therefore, one would expect incremental discounts to offer higher profits.

Table 3.1: Relative Profit of All-Unit Discounts to Incremental Discounts

	Heterogeneity (b)		
N	1.0	0.75	2.0
2	0.84	0.86	0.88
3	0.90	0.88	0.89
4	0.91	0.87	0.93
5	0.91	0.87	0.95
6	0.90	0.87	0.95

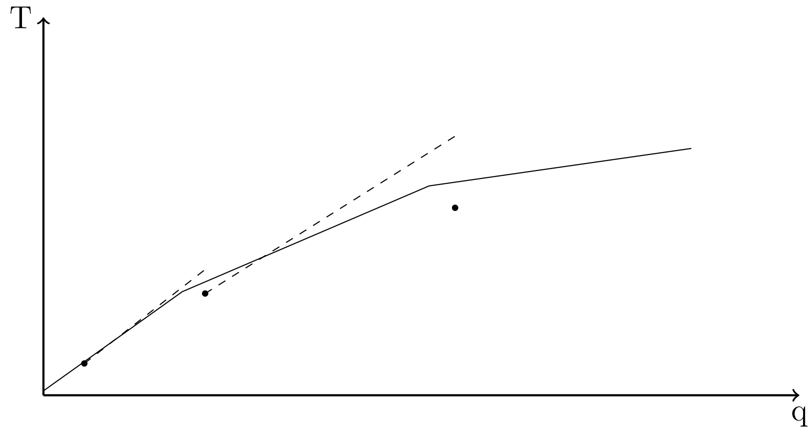
The relative value of all-unit discounts is higher when buyers are smaller and more similar, as shown in the third column of Table 3.1. As consumers are on average smaller, their purchase sizes will be smaller. The difference between the marginal prices on the incremental discount schedule and all-unit discount schedule are not as pronounced at smaller purchase sizes. Thus, when consumers make smaller purchases on average, the disparity between the two price schedules is reduced. Conversely, as consumers become larger on average the relative gain of offering incremental discounts is increased.

Finally, Figure 3.3 compares the total payments under the alternative discount schedules.[7] Note that the marginal prices are higher under the all unit discount schedule, as expected. Further, the total payment is lower under all-unit discounts at the cutoff quantities, indicated by the points on the graph. Finally, note that the maximum purchase size is lower under the all-unit discount schedule, reflecting the higher marginal prices.[8]

[7]For this figure, $N = 4$ and $b = 1.0$ corresponding to three prices and a uniform type distribution. The results are qualitatively similar under different parameterizations.

[8]Given this demand specification, it is optimal to set the upper bound on the bunching ranges equal to the lower bound on the next bunching range, i.e. $\overline{\theta}_k = \underline{\theta}_{k+1}$. This has two implications. First, the highest cutoff quantity is equal to the maximum purchase. Second, there are no purchases not at the cutoff quantities. This latter implication is interesting as it implies a discrete purchase size distribution despite

Figure 3.3: Comparison of Payments Under All-Unit and Incremental Discounts



3.5 Empirical Content of All Unit Discounts

Beyond the theoretical differences between all unit and incremental discounts, it is important to consider the observable implications of each type of discount. Data generated from each type of discount structure will differ in important ways. This section highlights those differences.

Consider a dataset that consists of a price schedule and purchase quantities. Given the price schedule and quantities, purchase prices can be derived depending on the type of discount schedule. Clearly, knowing the style of discounts is important. Quantities translate into different marginal prices and total payments depending on whether they are all-unit discounts or incremental discounts. For this discussion, assume the researcher knows the type of discounts implemented.

Recall that the major distinguishing feature of all-unit discounts is bunching in pur-

a continuous type distribution. It is unclear if this is a general property of the model or of the particular demand functional form.

chase quantities. This bunching is absent from incremental discount schedules. That is, given a set of purchase quantities, there should be no mass points in the distribution of quantity for incremental discounts while there should be several mass points for all-unit discounts.

Such bunching may present problems for estimation. Given a price schedule, purchase quantities, and total payments (either observed or derived), one could implement a version of the nonparametric estimation technique developed in (Ekeland, Heckman and Nesheim 2004).[9] However, their approach relies on identifying the buyer heterogeneity distribution from the purchase distribution. Such an approach may not be feasible given the bunching in the quantity distribution. That is, the quantity distribution has mass points while the buyer heterogeneity distribution does not, making nonparametric estimation more challenging.

3.6 Conclusion

The implementation of quantity discounts has important implications for what the quantity discounts look like and equilibrium behavior. This paper compares the two most common forms of quantity discounts: all-unit discounts and incremental discounts. All-unit discounts are primarily characterized by discontinuous total payment function where total payments may decline as purchases increase. This results in a bunching of purchases at the cutoff quantities. Both these features are important for equilibrium behavior and absent in an incremental discount schedule. Further, all-unit discount schedules see higher marginal prices and lower profits for the seller. These differences in equilibrium behavior also have implications for estimating the model parameters from observable purchases and payments.

[9] This approach is similar in spirit to the one implemented in (Perrigne and Vuong 2010).

Appendices

Appendix A

Simple Example of Ambiguous Welfare Consequences

The welfare consequences of second degree price discrimination in an intermediate goods market are ambiguous. This appendix details a simple model highlighting this fact. Considering a simple model also provides intuition for what factors determine whether allowing price discrimination increases or decreases total welfare.

Consider a single product monopolist wholesaler. The wholesaler has a marginal cost c of producing the good. The wholesaler sells the good as an input to a large number of retailers. Each retailer is a local monopolist who sells the single good and cannot store inventory.[1] Each retailer i faces local demand:

$$Q = \lambda_i A p_i^{-\epsilon}\,, \tag{A.1}$$

where

$$\lambda_i = \begin{cases} \lambda_h \text{ with probability } \alpha \\ \lambda_l \text{ with probability } 1-\alpha \end{cases}\,, \tag{A.2}$$

and $\lambda_l < \lambda_h$. That is, each retailer faces either high or low local demand for the product. Proportion α of the stores have high local demand. This definition of final consumer demand satisfies the single crossing property. For any price, high demand markets with type λ_h will sell strictly more than low demand markets with type λ_l.

[1] The retailer's production technology transforms one unit of input into one unit of output.

The wholesaler does not know (or cannot act on the knowledge of) the type of any particular retailer. However, the wholesaler does know the distribution of types given by λ_h, λ_l, and α. Therefore, the profit maximizing choice of the wholesaler is to engage in second degree price discrimination and offer a series of nonlinear tariffs, one for each type characterized by a fixed fee F_i and a marginal price ρ_i, and allow retailers to self-select among those options. The wholesaler must account for the high type retailers "pretending" to be the low type retailers and so must choose the tariffs to induce sorting among the retailers.

Formally, the wholesaler's problem is given by:

$$\max_{F_h,\rho_h,F_l,\rho_l} \alpha \left[(\rho_h - c)(Ap^\star(\rho_h)^{-\epsilon})\lambda_h + F_h\right] + (1-\alpha)\left[(\rho_l - c)(Ap^\star(\rho_l)^{-\epsilon})\lambda_l + F_l\right] , \tag{A.3}$$

$$\text{s.t. } (p^\star(\rho_l) - \rho_l)(Ap^\star(\rho_l)^{-\epsilon})\lambda_l - F_l \geq 0 , \tag{A.4}$$

$$(p^\star(\rho_h) - \rho_h)(Ap^\star(\rho_h)^{-\epsilon})\lambda_h - F_h \geq (p^\star(\rho_l) - \rho_l)(Ap^\star(\rho_l)^{-\epsilon})\lambda_h - F_l , \tag{A.5}$$

where $p^\star(\rho_i)$ is the optimal retail price charged by retailer facing marginal cost ρ_i. The first constraint ensures that even the low demand retailers want to purchase from the wholesaler. The second constraint says that the high type retailer profit for selecting the tariff designed for the high type (given by F_h and ρ_h) must be at least as large as the profit from the high type selecting into the tariff designed for the low type (given by F_l and ρ_l). This so-called incentive compatibility constraint ensures that the retailers are sorted into the correct tariff in equilibrium.[2]

[2]There are two additional constraints to be considered that do not bind in equilibrium and so are not discussed. There is an analogous non-negative profit condition for the high type and an incentive compatibility condition for the low type to not want to select into the high type tariff.

In this simple model, the expression for the optimal retail price is given by:

$$p^\star(\rho) = \frac{\rho\epsilon}{\epsilon - 1}\,. \tag{A.6}$$

Noting that the constraints will bind in equilibrium and substituting in the expression for the optimal retail price, the fixed fees can be expressed in terms of the wholesale prices:

$$F_l^\star = \left(\frac{\rho_l\epsilon}{\epsilon-1} - \rho_l\right)\lambda_l A\left(\frac{\rho_l\epsilon}{\epsilon-1}\right)^{-\epsilon}, \tag{A.7}$$

$$F_h^\star = \left(\frac{\rho_h\epsilon}{\epsilon-1} - \rho_h\right)\lambda_h A\left(\frac{\rho_h\epsilon}{\epsilon-1}\right)^{-\epsilon} - \left(\frac{\rho_l\epsilon}{\epsilon-1} - \rho_l\right)\lambda_h A\left(\frac{\rho_l\epsilon}{\epsilon-1}\right)^{-\epsilon} + F_l^\star\,. \tag{A.8}$$

Substituting these into the wholesaler's problem ensures the constraints are satisfied and allows the problem to be expressed in terms of just the wholesale price variables:

$$\begin{aligned}\max_{\rho_h,\rho_l} \alpha \Bigg[&(\rho_h - c)\lambda_h A\left(\frac{\rho_h\epsilon}{\epsilon-1}\right)^{-\epsilon} + \left(\frac{\rho_h\epsilon}{\epsilon-1} - \rho_h\right)\lambda_h A\left(\frac{\rho_h\epsilon}{\epsilon-1}\right)^{-\epsilon} \\ &- \left(\frac{\rho_l\epsilon}{\epsilon-1} - \rho_l\right)\lambda_h A\left(\frac{\rho_l\epsilon}{\epsilon-1}\right)^{-\epsilon} + \left(\frac{\rho_l\epsilon}{\epsilon-1} - \rho_l\right)\lambda_l A\left(\frac{\rho_l\epsilon}{\epsilon-1}\right)^{-\epsilon}\Bigg] + \\ &(1-\alpha)\left[(\rho_l - c)\lambda_l A\left(\frac{\rho_l\epsilon}{\epsilon-1}\right)^{-\epsilon} + \left(\frac{\rho_l\epsilon}{\epsilon-1} - \rho_l\right)\lambda_l A\left(\frac{\rho_l\epsilon}{\epsilon-1}\right)^{-\epsilon}\right],\end{aligned} \tag{A.9}$$

Solving the first order conditions of the above problem gives expressions for the optimal wholesale prices:

$$\rho_l^\star = \frac{c\epsilon}{\epsilon - \frac{\alpha}{1-\alpha}\frac{\lambda_h - \lambda_l}{\lambda_l}}\,, \tag{A.10}$$

$$\rho_h^\star = c\,. \tag{A.11}$$

The high demand markets are charged the efficient price. Low demand markets are charged a markup over marginal cost that depends on the relative frequency of each type and the relative size of big and small markets.

To find the optimal price when price discrimination is not allowed, the above problem can be resolved with a single price (and fixed fee) ignoring the incentive compatibility constraint. The optimal fixed fee and price without price discrimination are given by:

$$F^\star = \left(\frac{\rho\epsilon}{\epsilon-1} - \rho\right)\lambda_l A \left(\frac{\rho\epsilon}{\epsilon-1}\right)^{-\epsilon}, \tag{A.12}$$

$$\rho^\star = \frac{c\epsilon}{\epsilon-1}\frac{(\epsilon-1)(\alpha\lambda_h + (1-\alpha)\lambda_l)}{(\epsilon-1)(\alpha\lambda_h + (1-\alpha)\lambda_l) + \lambda_l}. \tag{A.13}$$

The final piece to calculating total welfare is consumer surplus. Consumer surplus in a particular market is:

$$CS_i(p) = \frac{A\lambda_i}{\epsilon-1}p_i^{-\epsilon+1}. \tag{A.14}$$

Total consumer surplus is thus given by the weighted sum:

$$CS = \alpha CS_h(p^\star(\rho_h^\star)) + (1-\alpha)CS_l(p^\star(\rho_l^\star)) \tag{A.15}$$

Given these solutions, determining whether second degree price discrimination by the wholesaler is welfare improving or not is a matter of comparing welfare under each solution.

Clearly the welfare calculation will depend on the values selected for the parameters $A, \epsilon, c, \lambda_l, \lambda_h$, and α. To illustrate that the welfare consequences of allowing second degree price discrimination by the wholesaler are ambiguous, it suffices to compare two particular parameterizations of the model. Further, comparing these two sets of parameters will give some insight into what causes the ambiguity. Table A.1 gives the two parameterizations under consideration.

Table A.2 presents the equilibrium values for the wholesaler choice variables under each parameterization. With relatively fewer high demand retailers, the low demand retailers

Table A.1: Parameter Values

Parameter	Scenario: Few Big	Scenario: More Big
Wholesaler Marginal Cost (c)	1.000	1.000
Demand Scale (A)	1.000	1.000
Elasticity (ϵ)	4.000	4.000
High Type (λ_h)	1.000	1.000
Low Type (λ_l)	0.250	0.250
Probability of High Type (α)	0.150	0.300

experience lower marginal prices. Correspondingly, the fixed fee is higher. Note that in both cases, high type retailers pay more for a marginal unit and low type retailers pay less when price discrimination is banned.

Table A.2: Equilibrium Values

Variable	Scenario: Few Big	Scenario: More Big
High Type Fixed Fee (F_h)	0.054	0.081
High Type Marginal Price (ρ_h)	1.000	1.000
Low Type Fixed Fee (F_l)	0.017	0.008
Low Type Marginal Price (ρ_l)	1.153	1.474
Uniform Fixed Fee (F)	0.021	0.018
Uniform Marginal Price (ρ)	1.084	1.134

Finally, Table A.3 presents the percent change in each welfare measure. Changes are presented for the high demand markets, low demand markets, and total market separately to help diagnose the source of the difference in the two scenarios.

As expected, the total change in wholesaler profit is negative in both scenarios. This

Table A.3: Welfare Percent Changes

	Scenario					
	Few Big			More Big		
Welfare Measure	High	Low	Total	High	Low	Total
Wholesaler Profit	-25.73	6.01	-2.96	-45.84	51.29	-14.89
Retailer Profit	20.16	0.00	20.16	119.28	0.00	119.28
Consumer Surplus	-21.52	20.16	-1.49	-31.49	119.28	-8.24
Total			0.15			-0.60

reflects the fact that the wholesaler can always do at least as well as the uniform price when they are allowed to discriminate.

Retailer profit also increases in both scenarios. As the table indicates, this is driven entirely by the high demand retailers. Marginal input costs increase for the high demand retailers, decreasing the amount of the good they sell, but the related fixed declines enough that total costs go down and profits go up. These two facts appear stable over a wide variety of parameterizations tested.

Overall consumer surplus declines in both scenarios. However, the result of removing quantity discounts is different depending on which market is considered. The changes mirror those seen in the retail tier. The increase in marginal input costs to the high demand retailers increases the resulting retail price and decreases consumer surplus. The opposite effect happens in the low type markets. Marginal input costs decline and so the retail price in those markets declines, increasing consumer surplus. The relative frequency of the two market types determines if the overall change in consumer surplus is positive or negative.

The sign of the total welfare change differs between the two scenarios. This is driven entirely by the changes in retail profits. The sign of the welfare changes are the same in both scenarios. However, the changes are larger in the scenario with more large retailers. The gain to retailer profit in this scenario is large enough to outweigh the losses to the wholesaler and consumers.

Whether total welfare goes up or down depends on the relative frequency of each market type. Anecdotal testing indicates that the relative size of λ_h and λ_l also matters. A more general statement would be that the change in total welfare depends on the relative size of each market as determined by λ_h, λ_l, and α.

This simple model highlighted the fact that the welfare consequences of second degree price discrimination in an intermediate goods market are ambiguous. The simple model also made clear that the distribution of retailer heterogeneity (in this case heterogeneity in demand) is crucial to understanding the sign of the total welfare change.

Appendix B

Solving for the Optimal Price Schedule

Solving the wholesaler's problem for the optimal price schedule is a well-defined, but computationally intense problem. The computational challenges arise from two central components of the model: the optimal retailer price for each segment, and the distribution of retailer types. This appendix discusses these issues further and details the computational approach used in solving the problem. Throughout the appendix, the product subscript is suppressed.

Recall that the wholesaler's problem is given by:

$$\begin{aligned}\pi_w = &\sum_{k=1}^{N-1} \int_{\lambda_k}^{\lambda_{k+1}} [(\rho_k - c))s(p^\star(\rho_k))\lambda_r M]\, dF(\lambda_r) \\ &+ (1 - F(\lambda_k))\lambda_k M((p^\star(\rho_k) - \rho_k)s(p^\star(\rho_k)) - (p^\star(\rho_{k-1}) - \rho_{k-1})s(p^\star(\rho_{k-1})))\,.\end{aligned} \tag{B.1}$$

The wholesaler picks a vector of wholesale prices $(\rho_1, \ldots, \rho_{N-1})$ and type cutoffs $(\lambda_1, \ldots, \lambda_{N-1})$ corresponding to quantity thresholds where the price schedule changes. The first order condition describing the optimal wholesale price for segment k is:

$$\begin{aligned}&\left[(\rho_k - c)s'(p^\star(\rho_k))\frac{\partial p^\star}{\partial \rho_k} + s(p^\star(\rho_k))\right] \int_{\lambda_k}^{\lambda_{k+1}} \lambda_r dF(\lambda_r) \\ &\qquad + \left[(p^\star(\rho_k) - \rho_k)s'(p^\star(\rho_k))\frac{\partial p^\star}{\partial \rho_k} + s(p^\star(\rho_k))(\frac{\partial p^\star}{\partial \rho_k} - 1)\right] \\ &\qquad\qquad \times [\lambda_k(1 - F(\lambda_k)) - \lambda_{k+1}(1 - F(\lambda_{k+1}))] = 0\,,\end{aligned} \tag{B.2}$$

and the first order condition for the optimal cutoff λ_k is:

$$
\begin{aligned}
&-(\rho_k - c)s(p^\star(\rho_k))\lambda_k f(\lambda_k) + (\rho_{k-1} - c)s(p^\star(\rho_{k-1}))\lambda_k f(\lambda_k) \\
&+ [1 - F(\lambda_k) - \lambda_k f(\lambda_k)]\left[(p^\star(\rho_k) - \rho_k)s(p^\star(\rho_k)) - (p^\star(\rho_{k-1}) - \rho_{k-1})s(p^\star(\rho_{k-1}))\right] = 0\,. \quad \text{(B.3)}
\end{aligned}
$$

These first order conditions form a system of $2(N-1)$ equations that describe the optimal price schedule. In principle, this system of equations can be solved using standard techniques for solving nonlinear equations.

Complications arise because of the presence of the optimal retail price for segment k given by $p^\star(\rho_k)$. If the optimal retail price cannot be found analytically, this introduces a nested computational loop into each evaluation of the first order condition for each segment. That is, for each iteration in the solution method for the optimal price schedule (i.e. guess of $\rho_k,\ k = 1, \ldots, N-1$), a numerical method would need to be employed to find the set of retailer prices $p^\star(\rho_k),\ k = 1, \ldots, N-1$. In the model presented in the main body of the text, the optimal retail price is characterized by:

$$
(p^\star - \rho_k)s'(p^\star) + s(p^\star) = 0\,. \quad \text{(B.4)}
$$

Combined with the assumption that $s(p)$ is given by a logit model, this implies that there is no analytical solution for $p^\star(\rho_k)$. Thus, in the empirical model of the paper such nested numerical methods are required.

Using a numerical method to find $p^\star(\rho_k)$ introduces some additional numerical error in the calculation of the first order conditions. The numerical method used to find $p^\star(\rho_k)$ will be an approximation to the true value. Thus, sufficiently tight numerical tolerances on this inner loop must be used to ensure that numerical noise does not propagate through the rest of the solution.

The other computational concern that arises in solving for the optimal schedule is the presence of the integral in the ρ_k first order conditions. Unless the integral:

$$\int_{\lambda_k}^{\lambda_{k+1}} \lambda_r dF(\lambda_r) \tag{B.5}$$

has a closed form, an additional numerical method must be employed to evaluate the integral. There are a variety of methods to evaluate such integrals numerically. In practice, I use a method based on quadrature. Simulation is another common method, but is more demanding computationally. I implement a nested quadrature rule using seven nodes detailed on `www.sparsegrids.de`. This rule will approximate an eleventh degree polynomial exactly. The numerical evaluation of the integral increases the computational burden in proportion to the number of nodes in the quadrature rule employed. There is an intrinsic trade off between the number of nodes and the accuracy of the approximation. Using a more accurate rule in my solution algorithm does not substantially change the results.

In additional to more computational burden (in the form of additional function evaluations), the numerical evaluation of the integral results in some numerical noise. Inherently, numerically evaluating the integral is an approximation process. When solving the system for the optimal schedule, the numerical tolerance of the solution must account for the approximation error in the integral. Choosing a sufficiently accurate approximation for the integral minimizes the chance that this approximation error impacts the solution results.

To summarize, requiring numerical methods to find the optimal retailer price $p^\star(\rho_k)$ and to approximate the integral introduces two kinds of computational problems. First, additional function evaluations, whether to solve the inner optimization problem to find $p^\star(\rho_k)$ or to approximate the integral, slows down computation. Second, these numerical methods introduce noise into the solution. Tight tolerances on the nested solutions and good approximations minimize the impact on the overall solution, but the "outer" tolerance for the overall solution can be no smaller than the approximation error introduced by the lower level numerical methods.

Setting aside the computational concerns outlined above, there is still the question of how to actually solve the system of equations to get the optimal schedule. In practice, I use an optimization based approach using a variation on Newton's method. Maximizing wholesaler profit with respect to the prices ρ_k and cutoffs λ_k for $k = 1, \ldots, N-1$ is a well defined problem. The gradient is given by the first order conditions outlined above and the Hessian is readily calculated.

Most multidimensional numerical optimization methods are local in nature and require a starting guess to initialize the algorithm. The efficiency of the numerical algorithm and perhaps the ability to even converge to a solution depends on how close the initial guess is to the actual optimum. Poor starting guesses lead to slow convergence or no solution.

In practice, I tried several approaches for picking starting values before settling on a continuation-based approach. I considered an approach based on linearizing the model as well as using several thousand iterations of a global solution method as other methods for choosing starting values. In practice, these approaches appeared inferior in terms of convergence time and stability to the continuation based approach described here.

In the continuation based approach, I solve the sequence of optimal tariffs for $\tilde{N} =$

$2, \ldots, N$ and use the solution for $\tilde{N}$ as the starting values for $\tilde{N}+1$. For example, to find the optimal schedule with six prices ($N = 7$) requires finding the optimal tariff for one price, two prices, three prices, etc. The optimal price and cutoff for the single price solution are used as the starting point for the two price schedule. The solution for the two price schedule is then used as the starting point for the three price schedule and so on.

In practice, this continuation based approach serves to provide stable numerical solutions for many prices and for a wide range of wholesaler parameters (the marginal cost c and distribution of types F). The solution for single price option appears to be relatively insensitive to the starting guess. Thus, this chain of solutions can reliably calculate the optimal solution for any number of price (i.e. any N) without being affected by the starting guess. It is, however, computationally demanding as the optimization problem must be solved repeated for every intermediate number of prices between one and the desired number.

In conclusion, the solution algorithm for and N part tariff is given below. Starting with $\tilde{N} = 2$ and some initial guess at a single price and cutoff:

Step 1: Find optimal price schedule for $\tilde{N}$ using Newton's Method.

Step 2: If $\tilde{N} = N$, take result as optimal price schedule. Else, continue to Step 3.

Step 3: Take resulting optimal price schedule consisting of optimal prices $\rho_k^\star$ and optimal cutoffs $\lambda_k^\star$ and append two new values corresponding to $\rho_{\tilde{N}}$ and $\lambda_{\tilde{N}}$. In practice I set:

$$\rho_{\tilde{N}} = \frac{\rho^\star_{\tilde{N}-1} + c}{2}$$
$$\lambda_{\tilde{N}} = \frac{\lambda^\star_{\tilde{N}-1} + 1}{2}$$

Step 4: Set new starting guess to $\rho_1^\star, \ldots, \rho_{\tilde{N}-1}^\star, \rho_{\tilde{N}}$ and $\lambda_1^\star, \ldots, \lambda_{\tilde{N}-1}^\star, \lambda_{\tilde{N}}$.

Step 5: Increase $\tilde{N}$ by one.

Step 6: Go to Step 1.

Bibliography

Aguirregabiria, Victor, Daniel Ershov, and Junichi Suzuki (2016) "Estimating the Effects of Deregulation in the Ontario Wine Retail Market," *Working Paper*.

Amornpetchkul, Thunyarat (2017) "Threshold discounts comparison: All-unit or incremental?" *Journal of Revenue and Pricing Management*, Vol. 16, pp. 265–294.

Armstrong, Mark (1996) "Multiproduct Nonlinear Pricing," *Econometrica*, Vol. 64, pp. 51–75.

Armstrong, Mark and John Vickers (2001) "Competitive Price Discrimination," *RAND Journal of Economics*, Vol. 32, pp. 579–605.

Ashenfelter, Orley C., Daniel S. Hosken, and Matthew C. Weinberg (2015) "Efficiencies brewed: pricing and consolidation in the US beer industry," *The RAND Journal of Economics*, Vol. 46, pp. 328–361.

Asker, John (2016) "Diagnosing Foreclosure due to Exclusive Dealing," *The Journal of Industrial Economics*, Vol. 64, pp. 375–410.

Bajari, Patrick, C. Lanier Benkard, and Jonathan Levin (2007) "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, Vol. 75, pp. 1331–1370.

Ben-Akiva, Moshe and Steven R. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, Cambridge, Mass: The MIT Press, 1st edition.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995) "Automobile Prices in Market Equilibrium," *Econometrica*, Vol. 63, pp. 841–90.

——— (2004) "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, Vol. 112, pp. 68–105.

Berry, Steven and Joel Waldfogel (2010) "Product Quality and Market Size*," *The Journal of Industrial Economics*, Vol. 58, pp. 1–31.

Besanko, David, Martin K. Perry, and Richard H. Spady (1990) "The Logit Model of Monopolistic Competition: Brand Diversity," *The Journal of Industrial Economics*, Vol. 38, pp. 397–415.

Busse, Meghan and Marc Rysman (2005) "Competition and Price Discrimination in Yellow Pages Advertising," *The RAND Journal of Economics*, Vol. 36, pp. 378–390.

Chu, Chenghuan Sean, Phillip Leslie, and Alan Sorensen (2011) "Bundle-Size Pricing as an Approximation to Mixed Bundling," *American Economic Review*, Vol. 101, pp. 263–303.

Conlon, Christopher T. and Julie Holland Mortimer (2013) "Efficiency and Foreclosure Effects of Vertical Rebates: Empirical Evidence,"Technical Report 19709, National Bureau of Economic Research, Inc.

Conlon, Christopher T. and Nirupama Rao (2015) "The Price of Liquor is Too Damn High: Alcohol Taxation and Market Structure," SSRN Scholarly Paper ID 2610118, Social Science Research Network, Rochester, NY.

Cooper, James C. and Joshua D. Wright (2012) "Alcohol, antitrust, and the 21st Amendment: An empirical examination of post and hold laws," *International Review of Law and Economics*, Vol. 32, pp. 379–392.

Copeland, Adam M. and Rodney J. Garratt (2015) "Nonlinear Pricing with Competition: The Market for Settling Payments," SSRN Scholarly Paper ID 2644601, Social Science Research Network, Rochester, NY.

Cosslett, Steven (1981) "Efficient Estimation of Discrete Choice Models," in Charles F. Manski and Daniel McFadden eds. *Structrual Analysis of Discrete Data with Econometric Applications*: MIT Press.

Crawford, Gregory S. and Ali Yurukoglu (2012) "The Welfare Effects of Bundling in Multichannel Television Markets," *American Economic Review*, Vol. 102, pp. 643–685.

DellaVigna, Stefano and Matthew Gentzkow (2017) "Uniform Pricing in US Retail Chains," Working Paper 23996, National Bureau of Economic Research.

Ekeland, Ivar, JamesJ. Heckman, and Lars Nesheim (2004) "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, Vol. 112, pp. S60–S109.

Grennan, Matthew (2013) "Price Discrimination and Bargaining: Empirical Evidence from Medical Devices," *American Economic Review*, Vol. 103, pp. 145–177.

Ho, Kate and Ariel Pakes (2014) "Hospital Choices, Hospital Prices, and Financial Incentives to Physicians," *American Economic Review*, Vol. 104, pp. 3841–3884.

Holmes, Thomas J. (2011) "The Diffusion of Wal-Mart and Economies of Density," *Econometrica*, Vol. 79, pp. 253–302.

Imbens, Guido (1992) "An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling," *Econometrica*, Vol. 60, pp. 1187–1214.

Inderst, Roman and Greg Shaffer (2009) "Market Power, Price Discrimination, and Allocative Efficiency in Intermediate-Goods Markets," *The RAND Journal of Economics*, Vol. 40, pp. 658–672.

Katz, Michael L. (1983) "Non-Uniform Pricing, Output and Welfare under Monopoly," *The Review of Economic Studies*, Vol. 50, pp. 37–56.

——— (1987) "The Welfare Effects of Third-Degree Price Discrimination in Intermediate Good Markets," *The American Economic Review*, Vol. 77, pp. 154–167.

Kolay Sreya, Shaffer Greg, and Ordover Janusz A. (2004) "AllUnits Discounts in Retail Contracts," *Journal of Economics & Management Strategy*, Vol. 13, pp. 429–459.

Manski, Charles F. and Steven R. Lerman (1977) "The Estimation of Choice Probabilities from Choice Based Samples," *Econometrica*, Vol. 45, pp. 1977–1988.

Manski, Charles F. and Elie Tamer (2002) "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, Vol. 70, pp. 519–546.

Maskin, Eric and John Riley (1984) "Monopoly with Incomplete Information," *The RAND Journal of Economics*, Vol. 15, pp. 171–196.

McFadden, Daniel (1977) "Modelling the Choice of Residential Location," Cowles Foundation Discussion Paper 477, Cowles Foundation for Research in Economics, Yale University.

McManus, Brian (2007) "Nonlinear pricing in an oligopoly market: the case of specialty coffee," *The RAND Journal of Economics*, Vol. 38, pp. 512–532.

Miravete, Eugenio J., Katja Seim, and Jeff Thurk (2017) "Market Power and the Laer Curve."

Mussa, Michael and Sherwin Rosen (1978) "Monopoly and product quality," *Journal of Economic Theory*, Vol. 18, pp. 301–317.

Nevo, Aviv (2001) "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, Vol. 69, pp. 307–342.

Nevo, Aviv, John L. Turner, and Jonathan W. Williams (2016) "Usage-Based Pricing and Demand for Residential Broadband," *Econometrica*, Vol. 84, pp. 411–443.

O'Brien, Daniel P. and Greg Shaffer (1994) "The Welfare Effects of Forbidding Discriminatory Discounts: A Secondary Line Analysis of Robinson-Patman," *The Journal of Law, Economics, and Organization*, Vol. 10, pp. 296–318.

Pakes, A., J. Porter, Kate Ho, and Joy Ishii (2015) "Moment Inequalities and Their Application," *Econometrica*, Vol. 83, pp. 315–334.

Perrigne, Isabelle and Quang Vuong (2010) "Nonlinear Pricing in Yellow Pages."

Petrin, Amil and Kenneth Train (2010) "A Control Function Approach to Endogeneity in Consumer Choice Models," *Journal of Marketing Research*, Vol. 47, pp. 3–13.

Quan, Thomas W. and Kevin R. Williams (2016) "Product Variety, Across-Market Demand Heterogeneity, and the Value of Online Retail," SSRN Scholarly Paper ID 2871513, Social Science Research Network, Rochester, NY.

Rochet, Jean-Charles and Lars A. Stole (2002) "Nonlinear Pricing with Random Participation," *The Review of Economic Studies*, Vol. 69, pp. 277–311.

Saffer, Henry and Markus Gehrsitz (2016) "The effect of post-and-hold laws on alcohol consumption," *International Review of Law and Economics*, Vol. 45, pp. 63–71.

Seim, Katja and Joel Waldfogel (2013) "Public Monopoly and Economic Efficiency: Evidence from the Pennsylvania Liquor Control Board's Entry Decisions," *American Economic Review*, Vol. 103, pp. 831–862.

Shum, Matthew (2004) "Does Advertising Overcome Brand Loyalty? Evidence from the Breakfast-Cereals Market," *Journal of Economics & Management Strategy*, Vol. 13, pp. 241–272.

Spence, Michael (1973) "Job Market Signaling," *The Quarterly Journal of Economics*, Vol. 87, pp. 355–374.

——— (1977) "Nonlinear prices and welfare," *Journal of Public Economics*, Vol. 8, pp. 1–18.

Stole, Lars A. (1995) "Nonlinear Pricing and Oligopoly," *Journal of Economics & Management Strategy*, Vol. 4, pp. 529–562.

Tirole, Jean (1988) *The Theory of Industrial Organization*, Cambridge, Mass: The MIT Press, 1st edition.

Train, Kenneth E. (2009) *Discrete Choice Methods with Simulation*, Cambridge ; New York: Cambridge University Press, 2nd edition.

Villas-Boas, Sofia Berto (2009) "An empirical investigation of the welfare effects of banning wholesale price discrimination," *The RAND Journal of Economics*, Vol. 40, pp. 20–46.

Wilson, Robert B. (1993) *Nonlinear pricing*, New York: Oxford University Press.